ORIGINAL RESEARCH

The ForEx++ based decision tree ensemble approach for robust detection of Parkinson's disease

Moumita Pramanik[1] · Ratika Pradhan[2] · Parvati Nandy[3] · Akash Kumar Bhoi[4,5,6,7] · Paolo Barsocchi[5]

Abstract

The progressive reduction of dopaminergic neurons in the human brain, especially at the substantia nigra is one of the principal causes of Parkinson's Disease (PD). Voice alteration is one of the earliest symptoms found in PD patients. Therefore, the impaired PD subjects' acoustic voice signal plays a crucial role in detecting the presence of Parkinson's. This manuscript presents four distinct decision tree ensemble methods of PD detection on a trailblazing ForEx++ rule-based framework. The Systematically Developed Forest (SysFor) and a Penalizing Attributes Decision Forest (ForestPA) ensemble approaches has been used for PD detection. The proposed detection schemes efficiently identify positive subjects using primary voice signal features, viz., baseline, vocal fold, and time–frequency. A novel feature selection scheme termed Feature Ranking to Feature Selection (FRFS) has also been proposed to combine filter and wrapper strategies. The proposed FRFS scheme encompasses Gel's normality test to rank and selects outstanding features from baseline, time–frequency, and vocal fold feature groups. The SysFor and ForestPA decision forests underneath the ForEx++ rule-based framework on both FRFS feature ranking and subset selection represents Parkinson's detection approaches, which expedite a better overall impact on segregating PD from control subjects. It has been observed that the ForestPA decision forest in the ForEx++ framework on FRFS ranked features proved to be a robust Parkinson's detection scheme. The proposed models deliver the highest accuracy of 94.12% and a lowest mean absolute error of 0.25, resulting in an Area Under Curve (AUC) value of 0.97.

Keywords Ensemble Parkinson disease detection · ForestPA · ForEx · Machine learning · Missing value imputation · Normality based feature selection · Parkinson detection · SysFor

✉ Moumita Pramanik
moumita.p@smit.smu.edu.in

✉ Akash Kumar Bhoi
akashkrbhoi@gmail.com

✉ Paolo Barsocchi
paolo.barsocchi@isti.cnr.it

Ratika Pradhan
ratika.sist@gmail.com

Parvati Nandy
parvati.n@smims.smu.edu.in

[1] Department of Computer Applications, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar, Sikkim 737136, India

[2] Department of Computer Engineering, Sikkim Institute of Science and Technology, South Sikkim, Jorethang, India

[3] Department of Medicine, Sikkim Manipal Institute of Medical Sciences, Sikkim Manipal University, Gangtok, Sikkim 737102, India

[4] Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar, Sikkim 737136, India

[5] Institute of Information Science and Technologies, National Research Council, 56124 Pisa, Italy

[6] KIET Group of Institutions, Delhi-NCR, Ghaziabad 201206, India

[7] AB-Tech eResearch (ABTeR), Sambalpur, Burla 768018, India

1 Introduction

The aged people suffered from a common neurodegenerative disorder often recognized as Parkinson's Disease. The disease associated with the human nervous system is often observed due to the anomalies of dopaminergic neurons located in the substantia nigra (Tuncer et al. 2020). Dopaminergic neurons help the brain to send signals to different parts of the body to carry out its corresponding function properly and supply correct articulation at the time of speech delivery. Common symptoms of PD are bradykinesia. The organ movement becomes very slow, tremor, rigidity, postural instability, walking/gait problems, decreased smell perception, sleep disturbances, and most importantly, variation in speech (Gómez-Vilda et al. 2017; Gupta et al. 2018; Tuncer and Dogan 2019; Aich et al. 2018; Tuncer et al. 2020). In PD, one side of the body part becomes distressed, and gradually dispersed to the other side of the body. The diagnosis of PD mostly depends on medical practitioners' motor tests when two out of three symptoms, namely akinesia (challenging to start a movement), rigidity, and rest tremor are observed. PD symptoms arise when the level of dopamine falls below 70% approximately (Jeancolas et al. 2017). As the disease is not reversible, preventive measures at an early stage and clinical intervention are necessary before the disease starts showing adverse symptoms.

According to Healthcare Research and Quality (Romexsoft 2017), doctors and clinicians are often prone to cognitive bias while diagnosing patients. Therefore, an intelligent system is needed to help clinicians to diagnose the patients efficiently. The system may employ Machine Learning (ML) (Anand et al. 2018) to eliminate these biases and help the medical practitioners to diagnose the disease with higher accuracy. ML makes healthcare smarter, especially in the context of a digital diagnosis of the disease. ML helps detect certain disease patterns using patients' electronic healthcare records and notify them of any anomalies. Over time, various researchers recommended different techniques and methods with different signals for detection of PD, such as voice signals (Pramanik et al. 2021a, b; Sakar et al. 2019; Mostafa et al. 2019; Gottapu and Dagli 2018), EEG (Yuvaraj et al. 2016), gait (Joshi et al. 2017; Zeng et al. 2016), biomarker (Bhat et al. 2018), MRI (Cigdem et al. 2018) and handwriting signals (Afonso et al. 2019; Rios-Urrego et al. 2019). Nevertheless, having various signals used for PD detection, the acoustic signal of the voice often considered for early detection of Parkinson's disease (Harel et al. 2004; Postuma et al. 2012; Rusz et al. 2016; Jeancolas et al. 2017). Therefore, vocal-based features attract the researchers to analyze PD in telemedicine (Ertuğrul et al. 2016; Bourouhou et al. 2016; Tuncer and Dogan 2019; Sakar et al. 2019; Jeancolas et al. 2017).

Hypokinetic dysarthria (Jeancolas et al. 2017) is a vocal injury found in the acoustics voice samples of Parkinson's patients, affecting multiple speech levels such as phonation, articulation, prosody, and resonance. Hypokinetic Dysarthria also influences the functions of laryngeal activity, articulatory and activities of respiratory muscles. Parkinson's patient often faces problem in speech formation; they experience disfluency and encounter problems in speech transformation as it becomes more monotonous. According to (Viswanathan et al. 2018; Jeancolas et al. 2017), the articulation of consonant and vowel delivery is gradually impaired, pitch intensity and phonation are unstable, and the tone becomes bumpy among Parkinson's subjects.

This article presents PD detection mechanisms using ForEx++ rule-based framework on the top of SysFor and ForestPA decision forests. The baseline, vocal fold along time–frequency feature groups are considered for PD detection. The voice feature groups were ranked and relevant features are acquired using a novel feature selection scheme based on Gel's directed test of normality. The proposed feature selection scheme is based on a Goodness of Fit (GF) test proposed by Gel et al. (2007), which facilitates designing the scheme both as the filter and wrapper-based approach. It is designed to rank (a filter method) the features based on their standard normal distribution and select a subset (a wrapper method) of normally distributed vocal features based on a critical value boundary. The proposed feature selection scheme can handle missing value instances before the feature selection procedure; thus, it proved to be a robust feature selection algorithm compared to peer feature ranking and selection schemes.

The chronological summary of the contributions of this article is as follows.

- A new feature selection scheme termed Feature Ranking to Feature Selection (FRFS) via Directed Tests of Normality (FRFS-DTN) has been introduced using the concept of Gel's Goodness of Fit (GF) test to acquire normally distributed features.
- A missing values imputation approach using neighbor information has been proposed as the pre-processing component of the proposed FRFS-DTN method; thus, making the feature selection method robust even in presence of outliers and missing values in the underlying data.
- The ranked features and selected feature subsets are used separately with SysFor and ForestPA to design four distinct PD detection mechanisms.
- Further, the ForEx++ rule-based framework has been introduced to expedite the detection result by selecting decision forest rules that are comparatively more accurate.

- The optimum acoustic features are proposed through the detection model to the research community to design the state-of-the-art PD detection methods.

The organisation of the article is as follows Sect. 2 narrates the activities of related works. Section 3 illustrates a schematic representation of the proposed PD detection model along with the datasets used, data preprocessing and feature selection technique. The results and discussion have been outlined in Sect. 4, followed by a conclusion and future research direction is mentioned in Sect. 5.

2 Related works

Machine learning-based Parkinson's detection mechanisms using acoustic signals has been explored extensively in the recent past. Most modern machine learning-based PD detection techniques employ feature selection techniques at the pre-processing stage to select suitable features. Smekal et al. (2015) suggested a Parkinson's finding approach utilizing Empirical Mode Decomposition (EMD) and Sequential Forward Feature Selection (SFFS) along with the Random Forest (RF) decision tree. EMD and SFFS have been used as feature selectors, which ultimately ascertained the optimum vowel features. The RF on SFFS proved to be an ideal PD detection scheme. Similarly, Mekyska et al. (2015) presented a voice-based PD detection approach using Random Forest, which was effective in disease progress prediction. The Random Forest PD detector reveals a sensitivity of 92.86% and specificity of 85.71%. Other than Random Forest, fuzzy entropy-based vocal features on Linear Discriminant Analysis (LDA) (Su and Chuang 2015) is also used to detect the presence of Parkinson's effectively among the subjects. Similarly, Bourouhou et al. (Bourouhou et al. 2016) compared different classification methods to detect Parkinson's disease with k-NN, Naïve Bayes, and SVM algorithm cross-validation. Though the SVM-based detection reveals magnificent performance, it is silent about a feature selection scheme. Employing a vibrant feature selection scheme could make the system more efficient. Similarly, four feature selection algorithms and six classifiers are evaluated to design a Parkinson's prediction engine (Cantürk and Karabiber 2016). The study shows that a total of 12 features generated by Relief feature selection, when trained on Back Propagation Neural Network (BPNN) consisting of twenty neurons, results in accuracy of 68.94%. The author also proved that feature selection from discriminative voice features is required to improve detection scores. In Orozco-Arroyave et al. (2016), the authors performed a rapid repetition of syllables of German, Spanish and Czech languages to segregate the presence of Parkinson's from control subjects.

Voice-based features, viz., phonation, articulation, and prosody play a lead role in identifying Parkinson's.

In (Chandrayan et al. 2017), PD detection is carried out based on dominant feature selection techniques where factor analysis feature selection played a crucial role in selecting dysphonia features from the speech signal. The author suggests an SVM-based detection on the feature subset that revealed an accuracy of 90%. Similarly, in (Jeancolas et al. 2017), twelve features from voice signal were generated using Mel-Frequency Cepstral Coefficients (MFCC), which has been fed into multi-dimensional Gaussian Mixture Models (GMM) for detecting positive PD patients. The author uses sustained vowels, fast syllables, reading, and free speech signals of PD patients and normal subjects. Their system shows an accuracy of about 91% for reading tasks. MFCC features are also extensively used along with Intrinsic Mode Functions (IMF) because of their ability to characterize Parkinson's dysphonia (Rueda and Krishnan 2017). The authors successfully achieved the quality of the audio signal with a visible distinction using MFCC features. The sustained vowel with sampling frequency at 8 kHz and quantization at a reduced resolution of 8-bit were used in the validation task. Dinesh et al. (Dinesh and He 2017) proposed filter-based feature selection techniques with many classifications, including Locally-Deep SVM (LDSVM), Decision Forests (DF), Logistic Regression (LR), Boosted Decision Tree (BDT), SVM, and Neural Networks (NN). On the other hand, a Random Least Squares Feedforward Network (RLSFN) binary classifier used on pathological and normative features predicts Parkinson subjects efficiently (Gómez-Vilda et al. 2017).

Similarly, an ensemble mechanism of classification have been used for Parkinson's detection (Fayyazifar and Samadiani 2017). In the pretext of the classification, the Genetic Algorithm plays a crucial role in selecting only six-voice features that have improved the classification accuracy to a great extent. A dual-stage Bayesian filter-based feature selection and classification approach for Parkinson's disease detection has been proposed using voice recording replications (Naranjo et al. 2017). One feature per feature group has been selected at the first stage, which is fed into a regularization-based classification approach based on Least Absolute Shrinkage and Selection Operator (LASSO) in the later stage. The dual-stage detector successfully classifies the Parkinson's and normal subjects with an accuracy of 86%, a sensitivity of 82.5%, and a specificity of 90%.

Gupta et al. (2018) used the optimized cuttlefish algorithm as a search strategy to ascertain the optimal subset of the feature on different types of sound recording and handwriting sample's dataset and decision tree. A non-linear decision tree and random forest classifier were used on two feature sets original feature sets and PCA feature sets to detect Parkinson's disease. In Viswanathan et al. (2018),

features extracted from sustained voiced consonant /m/ compared with the sustained phonation /a/ were considered where in SVM classifier model and the spearman correlation coefficient analysis was carried out to find the Unified Parkinson's Disease Rating Scale (UPDRS) motor score. A new expert system-based PD detection system proposed in Montaña et al. (2018) differentiates healthy subjects from people who have Parkinson's disease using Diadochokinesis tests. The system used temporal and spectral features extracted from the Voice Onset Time (VOT) segments of / ka/ syllables. The VOT algorithms was used to smoothen the amplitude envelope of the signal so that the classification algorithm can get meaningful distinctions about the subjects. The authors demonstrated that the novel VOT algorithm accurately estimates VOT boundaries for healthy and PD-affected subjects. An automated feature learning scheme was proposed to eliminate manual feature selection and intervention of experts (Wu et al. 2018). The system is based on a Mel-spectrogram of the audio signals of the subjects. The system employs derivative of Mel-spectrogram with respect to time for pre-processed audio signals. The extracted features are trained and tested through spherical k-means, which was the core of the detection process. Though the system shows adequate accuracy of detection, but the authors admitted that learning through MFCC has low clinical interpretability.

Sakar et al. (2019) analyses speech signal processing using multiple classification schemes. The Tunable Q-Factor Wavelet Transform (TQWT) and MFCC techniques extract required features from acoustic signals of vowels. Subsequently, the mRMR feature selection scheme generates the significant features passed to multiple classifiers. The predictions of the classifiers were combined with the ensemble learning approaches to come across a conclusive result. Parkinson's disease detection using L2-regularized logistic regression, random forest, and gradient boosted decision trees on voice samples (Tracy et al. 2019) has been proposed using 62 voice features. Their study shows that the input data is mostly skewed towards control patients, which was why the author's considered Recall, Precision, and Area Under Curve (AUC) as the parameters for evaluation. Multiple voice recordings by simultaneous sample and feature selection have also been conducted for the early diagnosis of Parkinson's disease (Ali et al. 2019a, b). The author employed a Leave-One-Subject-Out (LOSO) (Sakar et al. 2013) validation scheme on Neural Network, which proved to be a practical validation scheme for voice datasets having more than one sample per subject. Before the LOSO validation, their proposed method ranks features using a chi-square statistical model, searches for an optimal subset of the ranked features, and samples were selected iteratively. The phonation and speech were used by (Almeida et al. 2019) for detecting Parkinson's. For phonation, sustained vowels, and

speech, the pronunciation of sentences was used. The voice samples were captured using Acoustic Cardioid (AC) and Smartphone (SP). The captured samples are pre-processed, and adequate features were selected for the detection process. The authors used k-Nearest Neighbor (k-NN), Multi-Layer Perceptron (MLP), Optimum-Path Forest (OPF), and Support Vector Machine (SVM) as the detectors optimized using a hold-out technique. Voice features have also been extensively explored in an uncontrolled background condition (Braga et al. 2019). The Random Forest optimized with grid-search and learning curves was the basis for such PD detection.

In 2020, Tuncer et al. (2020) proposed a combination of 3 levels of Average Minimum Maximum (MAMs) tree used for pre-processing the voice data and Singular Value Decomposition (SVD) was used for feature extraction where 50 most distinctive features were selected using Relief feature selection techniques. Solana et al. (2020) proposed an approach of PD detection, which proved to behave improved accuracy by reducing selected vocal features. Out of the 754 features of the public dataset, the number of features selected for classification was in a range of 8 to 20 through Wrappers feature subset selection. Four classifiers k-NN, MLP, SVM, and Random Forest was the basis of the detection engine. Karan et al. (2020) proposed an SVM-based PD detection system where Intrinsic Mode Function Cepstral Coefficient (IMFCC) feature extraction is used to extract the most relevant feature for Parkinson's and Control patients classification. The authors validated their proposed approach through the two most widely used voice datasets. The IMFCC feature selection approach shows an improvement of 10–20% in accuracy compared to the standard acoustic and Mel-Frequency Cepstral Coefficient (MFCC) feature selection.

Braga et al. (2019) proposed a Parkinson's detection model using an optimized version of Support Vector Machine (SVM). SVM variation known as Radial Basis Function (RBF) has been optimized using an automated version of the grid-search technique. Their proposed grid search technique employs k-fold (k = 300) cross-validation to realize C = 1.15291 and γ = 0.00048. With the optimized RBF, Braga et al. approach detects the presence of Parkinson's with 92.38% accuracy. Similarly, Behroozi and Sami (2016) proposed a Parkinson's detection framework using multiple classifier ensembles. The majority voting techniques decided whether the subject suffers from Parkinson's or not. Ali et al. (2019a, b) suggested a neural network model encompassing genetic algorithms and linear discriminant analysis for Parkinson's detection. The approach was tested on a dataset that doesn't have control participants remains a drawback for the k-NN model. Agarwal et al. (2016) derived a model for Parkinson's detection using Extreme Learning Machine (ELM) on various acoustic features such as jitter, shimmer, pitch, period and harmonicity. Li et al. (2017)

discriminate Parkinson's subjects from controls using an ensemble framework combining the essence of random forest (RF), Support Vector Machines (SVM) and Extreme Learning Machine (ELM). Samples that have true separability are identified using Classification And Regression Tree (CART). The ensemble approach detects the presence of Parkinson's with up to 90% accuracy. Benba et al. (2016) proposed a PD detection model using k-NN with its different kernels (i.e., RBF, Linear, polynomial, and MLP). The k-NN model was tested on various voice samples where the approach reveals the highest detection accuracy of 82.5%. Berus et al. (et al. 2018) combined multiple artificial neural networks to form a PD detection model. Kendall's correlation coefficient acts as a feature selector of vocal signals of PD and control subjects. The neural network ensemble discriminates PD from control subjects with an impressive accuracy rate of 81.33%. A Multi-Edit-Nearest-Neighbor (MENN) algorithm for sample selection and Random forest (RF) with neural network ensembles proved to be an excellent Parkinson's detection engine (Zhang et al. 2016) under multiple acoustic features. Polat and Nour (2020) proposed a Parkinson's detection system using a novel one-against-all (OAA) sampling technique. The authors partition the data into five equal parts for Parkinson's and control subjects in this process. Logistic Regression, SVM and k-NN classifiers are combined for the training on the separate blocks of data. The classifiers combinations revealed a maximum detection accuracy of 89.46% while classifying PD subjects. A Mel scaled filter bank-based approach has been proposed for the early detection of Parkinson's disease (Upadhya et al. 2019). The MFCC features are used for detection purposes. The Radial Basis Function (RBF) plays a dramatic role in discriminating Parkinson's from healthy people. The RBF detector detects the presence of Parkinson's with more than 80% of accuracy. Adaptive Grey Wolf Optimization Algorithm (AGWOA) and Sparse Auto-Encoder (SAE) proved to be a sophisticated technique for vocal feature based Parkinson's detection (Xiong and Lu 2020). The SAE model on Correlation-based Feature Selection (CFS), Recursive Feature Elimination (RFE), minimum Redundancy Maximum Relevance (mRMR) has been tested through six supervised classification techniques. The Random Forest (RF) and Linear Discriminant Analysis (LDA) evolved as the best detector on the SAE. Naranjo et al. (2019) proposed a two stage classification approach to segregate Parkinson's from control subjects. The two stages classification approach has been designed to work specifically with replicated acoustic features. Common Principal Component (CPC) has been employed for dimensionality reduction. The CPC first extracted relevant features per acoustic feature group and secondly, the CPC extracted discriminated features from the combined features group. In this way, the two detection models are prepared, which proved effective

in diagnosing Parkinson's. Yaman et al. (2020) described a statistical pooling method for increasing the vocal features of the subjects. The Relief Feature section approach selects the most weighted features. Two different variations of SVM and k-NN have been used to derive the proposed Parkinson's detection scheme. According to the confusion matrix, both the k-NN and SVM variations show detection accuracy up to 91.25%. Recently, a two-stage whale optimization method for the classification of Parkinson's disease has been proposed (Ozturk and Unal 2020). Since most of the acoustic datasets contain replicated features, the subjects of Parkinson's and control classes are separated in the feature space. The instances of different classes are classified differently. The class separation approach detects the presence of Parkinson's effectively. A computer aided diagnosis method has been proposed using various kernels of SVM on acoustic signal features (Perez et al. 2016). The Radial Basis Function (RBF) revealed highest detection accuracy of 85.25% among the SVM kernels. Fuzzy neural network provides a potential alternating way of detecting Parkinson's in the acoustic signal (Guimarães, de Campos Souza, and Lughofer 2020). The model contains three layers, viz., Data Density Fuzzification, Fuzzy Rules and Artificial Neural Network. The combination of layers detects the presence of Parkinson's with 80.88% accuracy.

3 Materials and methods

The proposed PD detection scheme shown in Fig. 1 follows three distinct steps: pre-processing, Feature Ranking to Feature Selection via Directed Test of Normality (FRFS-DTN), and Parkinson's detection. At the pre-processing stage, the underlying data is aggregated, suitable feature groups are identified. In the FRFS-DTN stage, the features of each feature group are ranked, and finally, relevant features are selected. The feature selection provides the acoustic features after ranking according to the normality score. From the ranked features, the convenient features are selected based on *pValue* and upper ∝-percentile. Similarly, during the detection stage, the SysFor and ForestPA decision forest classifiers are used separately under ForEx++ rule-based framework. A reasonable number of decision trees on dynamically selected vocal features plays a vital role in Parkinson's detection. The proposed PD detector model works in iterative fashion. The number of decision trees in the forest with training and testing split instances is determined iteratively (Pramanik et al. 2021a, b). The dynamic selection of training and testing split provides a threshold point where the maximum detection accuracy has been achieved. The detailed steps of the Parkinson's detection model have been discussed in the following section describing the dataset

Fig. 1 The Parkinson's disease detection model
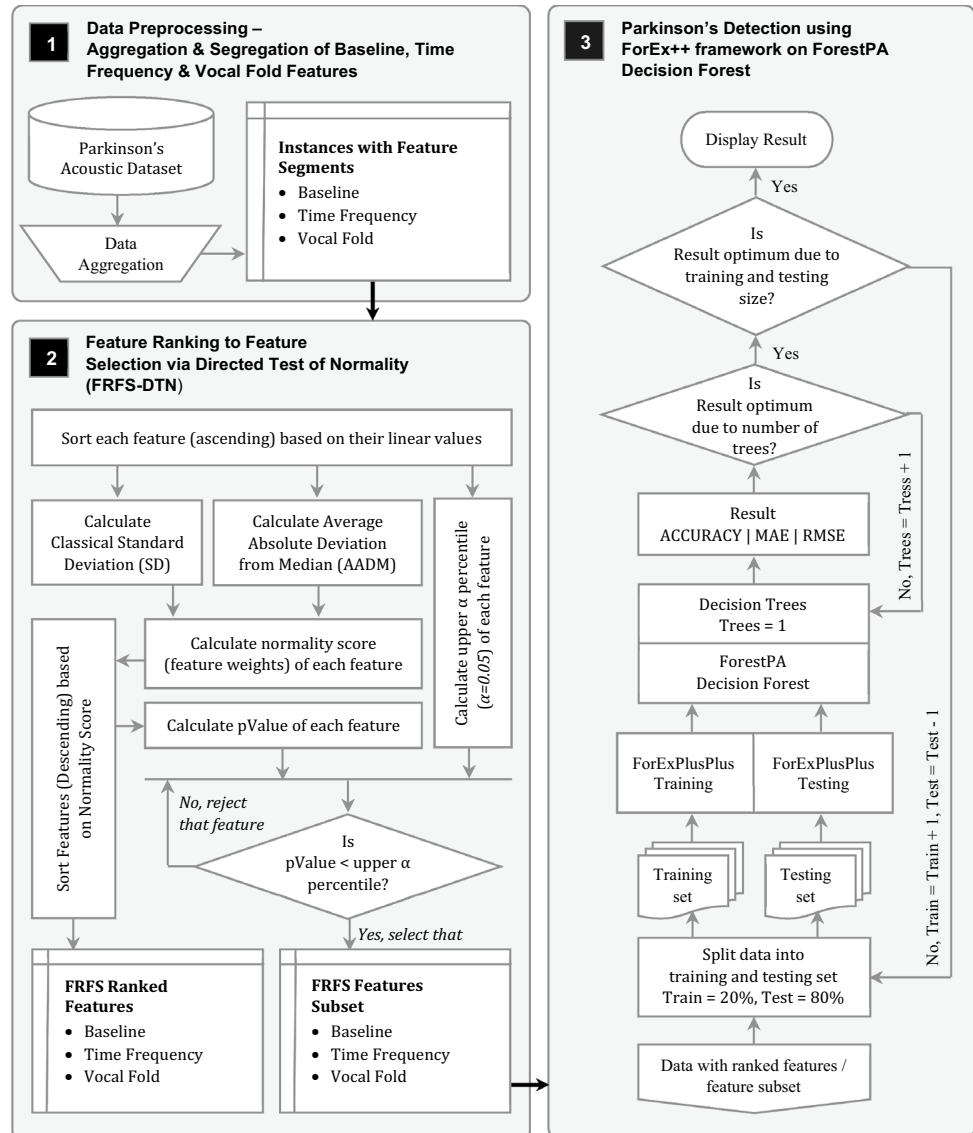
first, followed by feature ranking and selection procedure and then Parkinson's detection.

3.1 Dataset used

The dataset for analyzing Parkinson's disease lies in many formats. There are many datasets proposed so far for Parkinson's disease detection. These datasets have their advantages and disadvantages. Generally, the data of voice signals of PD patients are stored in terms of acoustic formats. However, the acoustic signals fall under three major groups- Baseline Features (BF), Time Frequency Features (TFF) and Vocal Fold Features (VFF). The BF are the combination of Detrended Fluctuation Analysis (DFA), Pitch Period Entropy (PPE), Harmonic to Noise Ratio (HNR), and various Jitter, Shimmer attributes of the acoustic signal. PPE is often calculated with different uncontrollable perplexing

effects of acoustic data in its voice frequencies. At first, pitch sequence is obtained and converted into the logarithmic scale. Roughness is analyzed by removing the linear temporal correlations with a standard linear whitening filter to produce the relative semitone various sequence. Next, a discrete probability distribution is constructed, and finally, the entropy of this probability distribution is calculated (Little et al. 2008; Edwards 2008). Detrended fluctuation analysis (DFA) is used to detect general voice disorders. It measures the degree of amplitude variation of the speech signal in a range of time scales and finds out the noise's similarity. The noise is often produced because of irregular airflow in the vocal fold (Little et al. 2008). The jitter parameter represents the frequency variation of voice cycles, whereas shimmer shows the amplitude variation of the sound wave (Farrús et al. 2007; Farrús and Hernando 2009). The frequency is generated by the number of sound waves formed

by the vocal cord's repetition during a particular time frame due to glottis' opening and closing. Fundamental frequency jitter and shimmer are often considered for vocal qualities analysis (Teixeira et al. 2013). Similarly, the TFF group is represented with various formant frequencies, intensity, and bandwidth. In this context, the intensity represents the vocal folds' volume due to the sub-glottis pressure's vividness (Mongia and Sharma 2014; Sakar et al. 2019). The formant frequencies are assessed using the frequency reaction of the vocal tract filter, whereas bandwidth ranges between b1 and b4 (Sakar et al. 2019). VFF consists of Glottis Quotient (GQ), glottal to Noise Excitation (GNE), Vocal Fold Excitation Ratio (VFER), and Empirical Mode Decomposition where the Glottis quotient gives the periodicity of glottis movement and GNE gives the quantities of turbulent noise. VFER provides the noise information generated by the vocal fold quivering by calculating non-linear energy and entropy. In contrast, EMD breaks a speech signal into a simple signal that applies basis functions adaptive in nature and calculates the energy/entropy values (Sakar et al. 2019).

In this article, three acoustic signal datasets have been considered for the evaluation of the proposed model. The first acoustic dataset has been prepared by Sakar et al. (2019). The dataset contains acoustic features of 252 persons in the age group of 33 to 87, where 64 healthy persons (33 male and 41 Female) and 188 PD persons (107 Male and 81 Female). The voice of subjects is recorded with a microphone with 44.1 kHz. A total of 754 features were extracted out of the sound files of subjects. The 754 features are spanned over six feature groups: BF, TFF, VFF, MFCC, Wavelet, and TQWT Features. Similarly, the second dataset prepared by Naranjo et al. (2016) is a balanced dataset consisting of acoustic signal features of 40 Parkinson's and 40 control subjects. Finally, the third acoustic dataset was proposed by Sakar et al. (2013). The dataset is spanned over separate training and testing sets of Parkinson's and control subjects. The training set comprises 40 subjects, out of which 20 are controls and 20 belong Parkinson's category. Similarly, the testing dataset holds acoustic information of 28 subjects suffering from Parkinson's. The acoustic information include pronunciation of sustained vowel /a/, /u/, /o/, short sentences and words. Since, other two datasets considered in this article holds acoustic information limited to the pronunciation of /a/ only; therefore, other acoustic information such as /u/, /o/, short sentences and words have been removed. Again, the training set contains only 40 instances, which may overfit or underfit the classifiers during the event of detection. Therefore, both the training and testing sets are merged to form a single dataset, where the training and testing splits have been determined dynamically. Both the Sakar et al. (Sakar et al. 2013, 2019) acoustic datasets are highly class imbalanced. On the other hand, Naranjo et al. (2016) dataset the most balanced due to equal number of Parkinson's and control subjects. So far as the number of features concern, Sakar et al. (2019) dataset contains 54, Sakar et al. (2013) contains 26 and the balanced Naranjo et al. (2016) dataset contains 18 features, respectively. The BF, TFF and VFF of these three datasets have been mentioned in Table 1.

Out of all the datasets mentioned in Table 1, the Sakar et al. (2019) dataset contains additional vocal information such as MFCC, Wavelet and TQWT, and BF, TFF and VFF. It has been observed that the most common parameters of BF, TFF and VFF used in the acoustic analysis are jitter, shimmer, and Harmonic to Noise Ratio (Teixeira et al. 2013). In a recent study conducted by us (Pramanik et al. 2021a, b), it is found that the BF, TFF and VFF show better detection results on decision trees than MFCC, Wavelet, and TQWT Features. Moreover, BF, TFF and VFF segments also provide the vital information about frequency, amplitude, noise, and vocal fold vibration. As the proposed work is based on decision forest, baseline, time frequency, and vocal fold feature groups are considered for Parkinson's detection over MFCC, Wavelet, and TQWT Features.

3.2 Data pre-processing

Sakar et al. (2013) dataset contains pronunciation of sustained vowel /a/, /u/, /o/, short sentences and words. The acoustic information of sustained vowel /a/ has been taken from the dataset for training and detection purposes. However, the Sakar et al. (2019) and Naranjo et al. (2016) datasets contain three recordings of sustained vowel /a/ for each subject. Applying machine learning techniques on the repeated instances will not provide meaningful information. Therefore, the average of the three recordings is calculated to get a unique record per participant. At the second step of data pre-processing the acoustic features of all the three datasets are ranked, and the suitable convenient features are selected according to their contribution towards Parkinson's detection. The ranking and selection are based on the normally distributed features of the datasets. Each feature of the datasets is assigned a normal distribution score, which decides whether the feature to be selected. The scores are calculated through a recent directed test of normality approach proposed by Gel et al. (2007). The Gel et al. approach has been modelled towards the feature selection; thus, Feature Ranking to Feature Selection (FRFS) via Directed Tests of Normality (FRFS-DTN). The FRFS-DTN proved to be a suitable feature selector for selecting normally distributed features. The method also acts as both filter and wrapper-based approach, i.e., it ranks and selects the prominent features. The proposed FRFS-DTN technique is novel because of its ability to handle missing values before the feature selection starts. The proposed FRFS-DTN feature

Table 1 The BF, TFF and VFF feature groups and number of features of Sakar et al. (2019), Naranjo et al. (2016); Sakar et al. (2013), datasets

Features segments considered for Parkinson's detection	Datasets		
	Sakar et al. (2019)	Naranjo et al. (2016)	Sakar et al. (2013)
	Number of features of the corresponding segment		
Baseline features			
Entropy of recurrence period density	1	1	
Detrended fluctuation (signal fractal scaling exponent)	1	1	
Entropy of pitch period (nonlinear measures of fundamental frequency variation)	1	1	
Harmonicity parameters (noise to tonal components in the voice)	2	5	3
Variants of jitter (measures of variation in fundamental frequency)	5	4	5
Fundamental frequency	5	–	
Variants of Shimmer (measures of variation in amplitude)	6	5	6
Time frequency features			
Voice intensity	3	–	
Bandwidth	4	–	
Formant frequencies	4	–	
Pitch parameters	–	–	5
Pulse parameters	–	–	4
Vocal fold features			
Glottis quotient	3	–	
Glottal to noise excitation	6	1	
Empirical mode decomposition	6	–	
Vocal fold excitation ratio	7	–	
Fraction of locally unvoiced frames	–	–	1
Number of voice breaks	–	–	1
Degree of voice breaks	–	–	1
Total number of features	54	18	26

ranking follows three major steps, i.e., handling missing values, feature ranking and feature subset selection.

3.2.1 Handling missing values

The proposed FRFS-DTN technique is suitable for data ranging from lower sample size (< 180) to higher sample size (≥ 180); thus, it proved to be scalable. However, a feature selection technique is robust if it can select the requisite features (generally features contributing maximum towards classification) even in the presence of outliers in the underlying data. Outliers include missing values, missing class information for a subject, and unstructured data. These are the main reason behind the classifier's performance degradation. However, there are many missing value imputation schemes in data science, but most of those methods predict the missing values with one or more supervised classifiers. Therefore, an effort has been made to design a missing value imputation algorithm that will act as a filter component for the FRFS-DTN feature selection scheme before the actual

ranking process starts. The missing value imputation technique has been detailed in Table 2.

The missing value imputation procedure takes the help of a traditional *k-means* clustering algorithm to impute the missing values. Here, NULL values for a cell have been considered as missing values. The imputation procedure has been conducted in the following four distinct steps. At first, the instances of data matrix M_D have been removed having no class information. It should be noted that the imputation method detailed in Table 2 employs *k-means* clustering in a supervised fashion. Therefore, no class information has no use, which needs to be eliminated from the data matrix permanently. After removing instances with no class information, the second stage involves removing instances where all cells have the missing values. Practically, a medical dataset cannot contain NULL values for all the cells of an instance. However, this precaution has been taken deliberately to make this algorithm extendible to other application domains. Therefore, if all the cell values of an instance have NULL, the instance itself is of no use, hence excluded from the

Table 2 Missing value imputation procedure for FRFS-DTN

INPUT

M_D = Data matrix having n number of instances and m number of features with random missing values and the mth position of M_D contains the class information.

OUTPUT

M_D = The imputed data matrix

PROCESS

for $i=1$ **to** n_c **do**
    Step 1. Create testing instance
    $T_s[i][j = 1 \dots \dots \dots m] := M_D^c[i][j = 1 \dots \dots \dots m]$,   where $M_D^c[k][j] \neq NULL$
    Step 2. Fetch rest of the rows as training instances
    for k:=1 **to** n_c
        $T_r[k][j = 1 \dots \dots \dots m] := M_D^c[k][j = 1 \dots \dots \dots m]$,   where $k \neq i, M_D^c[k][j] \neq NULL$
    end
    Step 3. Call *k-means* with training (T_r) and testing (T_s) instances with $k = |T_r|$ to obtain
    Predicted Cluster Number (PCN) for T_s
    $pcn := kmeans(T_r, T_s, k)$
    Step 4. Impute the ith row data of M_D^c with the data of predicted cluster
    $M_D^c[i][j = 1 \dots \dots \dots m] = T_r[pcn][j = 1 \dots \dots \dots m]$ where $M_D^c[i][j] = NULL$
end
return M_D

---

dataset. At the third stage of the process, attributes having all the cell values as NULL should be removed. A missing cell value cannot be imputed with another missing cell value of another instance. Once these three preliminary steps have been completed, each instance of the data matrix is processed for imputation. During the imputation step, instances are segregated and processed separately for each class label. There are two distinct class labels positive instances affected with Parkinson's and negative the control instances. For illustration, values containing diagnosis results of controls and subjects suffering from Parkinson's have different range values. Therefore, the segregation of data based on the class label is essential to achieve better-imputed values. So, in this stage, data of each class label has been processed separately. Each instance of a specific class label, where cell values are meant for imputation, has been treated as a testing instance, and the rest of the instances are treated as training instances. The training and testing instances have been prepared so that each NOT NULL cell of training instances has been considered matching NOT NULL cell values of the testing instance. Finally, the *k-means* cluster has been called with the desired training and testing instances. The general drawback of the *k-means* clustering is that the number of clusters formed needs to be defined before the actual clustering process starts. The drawback of defining the number of clusters beforehand actually turns out to be the strong point for the imputation. The number of clusters k has been fixed with

the number of training instances; thus, getting exactly one training instance in each cluster. During the testing process, the data NULL values of the testing instance rows of the data matrix can be imputed with the corresponding NOT NULL values of the cluster's training instance where the prediction of the testing instance falls. The entire process is repeated for each instance desired to be imputed of each class. After the successful imputation, the imputed matrix is passed to FRFS-DTN, presented in the next section.

3.2.2 Feature ranking to feature selection (FRFS) via directed tests of normality (FRFS-DTN)

The proposed FRFS-DTN procedure has been inspired by the concept of a directed test of normality. Almost all machine learning procedures rely on the illusion that the observed data are typically distributed. The training data suffers from (a) the curse of dimensions (b) the presence of skewness. The high dimensional data hampers the classifiers' performance, whereas the skewed data is the reason behind the generation of false alarms. Therefore, the normally distributed data with the correct dimensions is required to boost the performance of the classifiers.

Through various tests (Bonett and Seier 2002; D'Agostino 1986; Thadewald and Büning 2007) which helps to ensure that the underlying data is usually distributed, those approaches work perfectly well for small sample size data.

In 2007, Gel et al. (2007) presented a robust normality test for heavy-tailed samples. The approach generates a score for a set of data points, which helps to determine whether the data points are normally distributed or not. In this paper, the normality test inspired by Gel et al. approach has been adopted for optimum feature selection. The directed test of normality proposed by Gel et al. relies on the comparison of S_n – the classical standard deviation of data points and J_n – the average absolute deviation from the median of the data points. The ratio of S_n to J_n provides a legitimate score of normality. In the context of feature selection, the normality score has been calculated for each feature. The normality scores can be used for two purposes –

(a) Feature Ranking: Ranking features based on the normality score; thus, it acts as a filter-based feature selection.
(b) Feature Subset Selection: Selecting a subset of features from the ranked features like the wrapper-based feature selection.

Ranking individual features involves calculating each feature's score and arranging in descending order of their estimated score. The estimation of feature scores starts with calculating the average absolute deviation from the median of data points and is presented as –

$$AADM = \frac{C_1}{n} \sum_{k=1}^{n} |X_k - \tilde{X}| \tag{1}$$

where, $\tilde{X}$ = median of the data points, X_k= individual feature point, $k \leq n$, n = number of data points and $C_1 = \sqrt{\pi/2}$ a constant. Once the average absolute deviation from the median is calculated, the classical standard deviation of the data points X must be determined for varying sizes of data points. However, for $|X| >= 180$, the standard deviation can be calculated as –

$$SD = \sqrt{\frac{\sum_{k=1}^{n} X_k^2}{n} - \overline{X}^2} \tag{2}$$

Now the weight of the concerned feature can be estimated as the ratio of classical standard deviation and average absolute deviation from the feature's median. Therefore, the weight of the feature can be expressed as –

$$M_W = \frac{SD}{AADM} \tag{3}$$

The feature weight is the normal distribution score. More the weight, more the feature is normally distributed. Therefore, the features of the Parkinson dataset must be sorted horizontally according to the calculated feature weight. In this way, the feature having the highest weight secures the

first position with the highest rank. Similarly, the feature having the lowest weight is placed at the last position. The feature ranking procedure ranks the features in descending order of normality score. As a result, the first feature is highly normally distributed, and the last feature is a skewed representation of feature data. The challenge here is to identify the cut-off points of feature selection to select only the normally distributed features and drop the skewed features. The cut-off point for feature selection can be determined by calculating the *pValue* and upper $\propto$-percentile ($\propto= 0.05$) for all the feature vectors. The *pValue* can be calculated as –

$$pV = \frac{\sqrt{n}\big(M_W[i] - 1\big)}{C_2} \tag{4}$$

where, $M_W[i]$ represents the normality score of features i and $C_2 = \sqrt{\frac{(\pi-3)}{2}}$ is the asymptotic variance of $M_W[i]$. Any features having *pValue* more than or equal to the upper 0.05 percentile mean that the feature is skewed and dropped. All the features having *pValue* less than the upper 0.05 percentile is selected as ideal features.

3.3 Decision forest approaches

A decision tree is typically used for knowledge discovery by extracting logic rules from underlying data. Conversely, a decision forest extracts and manages rules with multiple decision trees ensemble. Two robust decision forests, viz; Systematically Developed Forest (SysFor) (Islam and Giggins 2011) and Penalizing Attribute Decision Forest (ForestPA) (Adnan and Islam 2017a), has been used as the primary detection component of the proposed PDS. However, decision forests believe to generate a vast number of logic rules during the training process. Therefore, SysFor and ForestPA are incorporated separately through a renowned knowledge discovery framework known as ForEx++ (Adnan and Islam 2017b). The ForEx++ framework is explicitly designed to handle decision forests to extract those comparatively more accurate, generalized, and concise rules than others.

The detection process through SysFor starts with identifying potential attributes, followed by a rule induction process. The induction process incorporates multiple C4.5 decision trees, where the voting mechanisms among C4.5 decision forests decides the presence of Parkinson's Disease. It should be noted that the proposed FRFS-DTN feature selection technique provides a list of suitable potential attributes. However, a more concise list of potential attributes can be realized through the attribute identification process of SysFor. In SysFor, the feature subset and ranked features generated by FRFS-DTN are scanned sequentially for feature refinement. The splitting point for SysFor prospective feature identification is represented as:

$$\frac{abs\left(P_j - P_k\right)}{|A_i|} - \beta > 0, \forall P_k \in P \tag{5}$$

where, P_j represent the splitting point (if available), P_k is represent the splitting point at position k. The (β) represents a threshold that shows the differences between significant and non-significant attributes. The attributes of gain ratio identify the essential attributes. The gain ratio of each attribute is estimated and maintained separately. The gain ratio of each attribute is used to arrange their position in decreasing order, where the favorable attributes are determined to have a gain ratio lesser than a goodness threshold value. The set of potential attributes identified are segregated horizontally into two groups according to their splitting points, which eventually helps to generate multiple C4.5 decision trees. The volume of trees present in the SysFor forest is calculated as:

$$T = \frac{\sum_{j=1}^{|S|} \left(\left|A_j\right| \times \left|D_j\right|\right)}{\sum_{j=1}^{|S|} D_j} \tag{6}$$

where, S = set of data segments, D = number of subjects, and A = useful attributes of S segments, respectively. The prediction result ascertained by T number of trees is placed for voting to realize the final decision about any subject, whether the subject is suffering Parkinson's or not.

Another potential decision forest was known as Decision Forest by Penalizing Attributes (ForestPA), has also been proposed for precise decision making with a facility for penalizing unworthy attributes at the tree building phase (Adnan and Islam 2017a). The Classification And Regression Trees (CART) are the basis of the ForestPA decision forests, where the trees are built upon the bootstrap samples S of the training instances D. The merit values of an attribute M_{A_i} is estimated as –

$$M_{A_i} = \left(C_{A_i} \times W_{A_i}\right) \tag{7}$$

where C_{A_i} specify the capacity of classification and W_{A_i} shows attribute weight (A). Subsequently, the weight of the potential attributes is increased iteratively from its default value 1. Finally, the depth of the decision tree determines the ultimate attributes' weight. The range of weights, WR_d for d tree-depth decreases the weight of non-potential attributes as:

$$WR_d = \begin{cases} [0.0000, e^{-1/d}] & if d = 1 \\ [e^{-1/(d-1)} + 0.0001, e^{-1/d}] & if d > 1 \end{cases} \tag{8}$$

The idea behind Eqs. (7) and (8) is to allocate weights to the lower-level nodes and penalizes higher-level nodes. The process eliminates the attributes acquiring the lowest weight of 0, thus increasing attributes weights dynamically at the root node. The increment value of weight WA_i^{++} of such attributes is evaluated as –

$$WA_i^{++} = \frac{1.0 - WA_i}{(h+1) - \ell} \tag{9}$$

where WA_i shows the weight of the attribute A_i estimated from a tree-level $\updownarrow$ with height h. The decision tree can predict the unlabelled instances using the dynamic weight allocation process and various class labels of different instances is determined through voting.

3.4 The ForEx++ framework

It is known that a decision forest has a better predictive ability than the standard decision tree. A typical decision forest, including but not limited to SysFor and ForestPA has certain deficiencies despite better prediction capability. The deficiency is the number of logic rules that the forest generates during the training process. Though the acoustic feature groups considered here are limited to 54 features, even with such limited features, a decision tree may generate vast numbers of logic rules that are difficult to manage and consume significant memory space and computational time. Adnan et al. (Adnan and Islam 2017b) proposed an independent dataset framework to eliminate irrelevant rules and identify those rules that are more concise, generalized, and accurate as a potential solution to this issue. It should be noted that the number of logic rules in the forest is already controlled by regulating the densities of decision trees in the forest (Fig. 1). However, further improvement can be possible by extracting the valuable rules from the remaining decision trees. The ForEx++ is designed explicitly for SysFor and ForestPA decision forests to extract proper consolidated rules which contribute most towards the detection process. For a set of rules $R = \{R_1, R_2, \ldots\ldots, R_z\}$, the consolidated rules of the underlying decision forest have been generated as:

$$R^{ForEx++} = \bigcup_{C_k \in C, \forall C_k} \left(R_{Avg,C_k}^{Accuracy} \cap R_{Avg,C_k}^{Coverage} \cap R_{Avg,C_k}^{Length}\right) \tag{10}$$

where, C_k is the class value for which the rule is being generated.

Mathematically, the $R_{Avg,C_k}^{Accuracy}, R_{Avg,C_k}^{Coverage}, R_{Avg,C_k}^{Length}$ are estimated as follows.

$$R_{Avg,C_k}^{Accuracy} = R_i : R_{i,C_k}^{Accuracy} \geq \frac{\sum_{j=1}^{\left|R_{C_k}\right|} R_{j,C_k}^{Accuracy}}{\left|R_{C_k}\right|} \tag{11}$$

$$R_{Avg,C_k}^{Coverage} = R_i : R_{i,C_k}^{Coverage} \geq \frac{\sum_{j=1}^{|R_{C_k}|} R_{j,C_k}^{Coverage}}{|R_{C_k}|} \qquad (12)$$

$$R_{Avg,C_k}^{Length} = R_i : R_{i,C_k}^{Length} \leq \frac{\sum_{j=1}^{|R_{C_k}|} R_{j,C_k}^{Length}}{|R_{C_k}|} \qquad (13)$$

Therefore, as a potential rule estimator, the ForEx++ framework first removes the identical rules from the set of rules R. Secondly, the $R_{Avg,C_k}^{Accuracy}$, $R_{Avg,C_k}^{Coverage}$, R_{Avg,C_k}^{Length} are estimated for each class through Eqs. (11, 12, and 13). Finally, ForEx++ consolidates all rules for all classes using Eq. (10). It should be noted that the FRFS feature ranking scheme ranks the features based on the normality score to achieve maximum detection accuracy. However, the order of attributes does not impact the number of rules generated by the decision forest. Therefore, the ForEx++ framework is essential to deduce the potential consolidated rules.

4 Results and discussion

The result and discussion have been carried out incrementally with several dimensions. At first, our proposed FRFS ranking scheme's performance outcome has been discussed briefly, along with other peer schemes. Secondly, the feature selection outcome of the FRFS scheme has been explored in detail. Finally, the proposed Parkinson's detection results on the ForEx++ framework using both ranking and selection of FRFS are elaborated briefly using various validation procedures. The validation procedure used here are Leave One Subject Out (LOSO) cross-validation, tenfold cross-validation and validation through training and testing split. Classification or detection approaches employ many parameters for understanding the capability of the classifiers. However, the applicability of performance measurement parameters is different for different application scenarios. For instance, the detection accuracy is suitable for measuring the ability of the detector to detect an incoming instance in almost all areas of data sciences. However, in the medical sciences, sensitivity and specificity are the measures helpful in evaluating the classifier's performance for detecting negative and positive subjects. The trade-off between sensitivity and specificity can be well understood by Area Under Curve (AUC) (Kumar and Indrayan 2011). In this section, the proposed models are explored using a variety of performance measures, i.e., Accuracy (ACC), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), False Negative Rate (FNR), False Positive Rate (FPR), Specificity (SPE), Sensitivity (SEN),

Area Under Curve (AUC) and Precision-Recall Curve (PRC).

4.1 Feature ranking to feature selection (FRFS)

The proposed FRFS-DTN feature ranking scheme inspired by the directed normality test has been designed for varying densities of data points. The standard deviation has been estimated to obtain each feature's normality score. Finally, the features are arranged in descending order of their normality score. The feature having the highest normality score will come; first, the feature having the second-highest score will come second, and so on. Arranging features based on normality scores is believed to improve the detection accuracy of the supervised classifiers. The proposed ranking of features has been validated through the vocal features of Sakar et al. (2019) (Sakar et al. 2019) dataset. The ranked acoustic features have also been passed to three popular supervised classifiers Naïve Bayes, OneR, and C4.5 decision tree, and three decision forest schemes SysFor, Random Forest, and ForestPA. The tenfold cross-validation has been conducted for observing classifiers' performance. Moreover, the classification has been undertaken with a change in feature size, which provides scope to understand the performance improvement of the classifiers with ranked features. It is observed that the classifiers' performance reaches the peak with a few features and gradually decreases with increased features count. The result of proposed FRFS-DTN scheme has also been compared with five other robust feature ranking schemes, viz., Correlation-based Feature Selection (CFS) (Roffo and Melzi 2017), FISHER (Gu et al. 2012), Feature Selection with Adaptive Structure Learning (FSASL) (Du and Shen 2015), Feature Selection via Concave Minimization (FSV) (Bradley and Mangasarian 1998) and Infinite Latent Feature Selection (ILFS) (Roffo et al. 2017). Figures 2, 3, 4 represents our proposed FRFS-DTN feature ranking scheme's performance with other peer schemes. Figure 2a, b, c, represents the detection accuracy of Naïve Bayes, C45 Decision tree, and OneR classifier, and Fig. 2d, e f presents SysFor, ForestPA and Random Forest's performance on BF segment respectively. Similarly, for VFF groups, Fig. 3a, b, crepresents detection accuracy of Naïve Bayes, C45 Decision tree and OneR classifier and Fig. 3d, e, f presents the performance of SysFor, ForestPA, and Random Forest. Finally, for TFF, Fig. 4a, b, c, represents detection accuracy of Naïve Bayes, C45 Decision tree, and OneR classifier, and Fig. 4d, e, f represents SysFor, ForestPA's performance Random Forest classifiers.

It can be seen from Fig. 2 that the proposed FRFS approach under BF segment shows superior performance than the peer CFS, FISHER, FSASL, FSV, and ILFS feature selections. The FRFS scheme reveals maximum detection accuracy for Naïve Bayes, C4.5 decision tree, and OneR
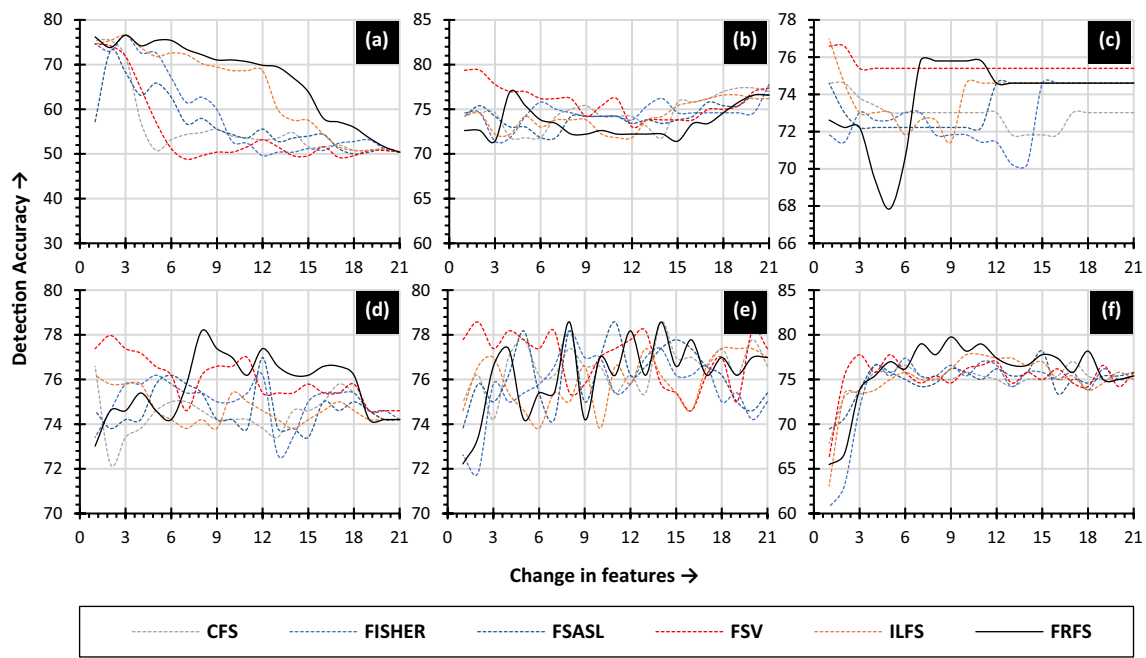
Fig. 2 Classification accuracy of the FRFS-DTN and other peer feature ranking schemes through **a** Naïve Bayes **b** C4.5 **c** OneR **d** SysFor **e** ForestPA **f** Random Forest classifiers on baseline features of Sakar et al. (2019) dataset
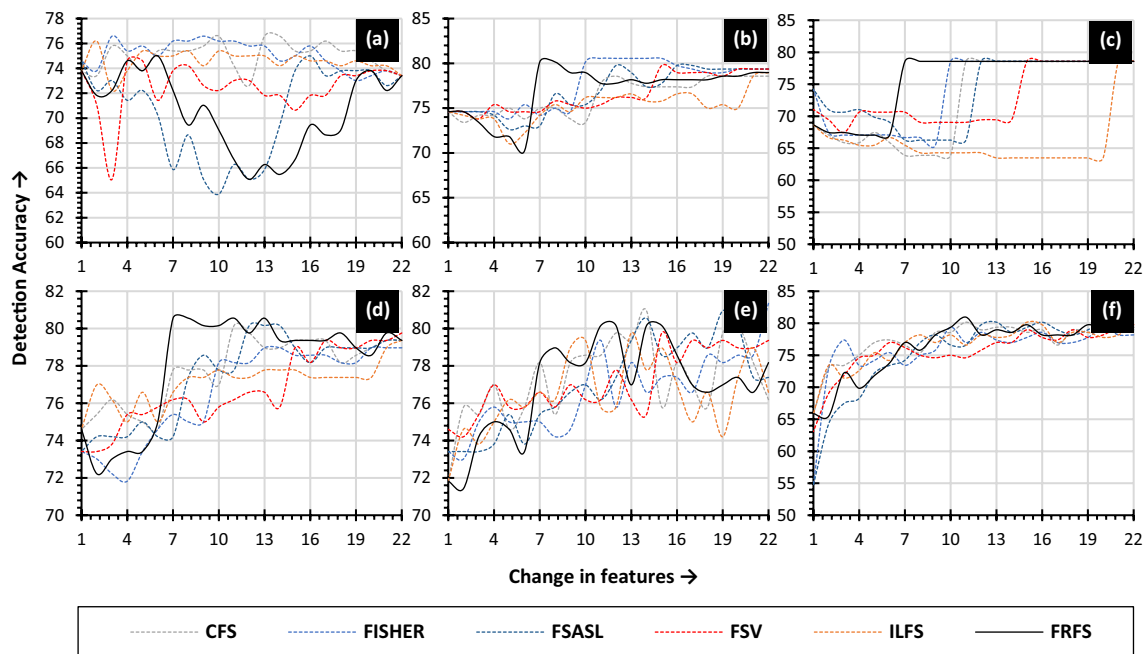

Fig. 3 Classification accuracy of the FRFS-DTN and other peer feature ranking schemes through **a** Naïve Bayes **b** C4.5 **c** OneR **d** SysFor **e** ForestPA **f** Random Forest classifiers on vocal fold features of Sakar et al. (2019) dataset

with just a few features ($m < 7$) in hand. In the case of Naïve Bayes, only the ILFS shows at par performance with the proposed approach. Nevertheless, the ranked 21 features prove that ranking is the better approach, as the maximum detection accuracy is observed with just little number of features in hand. The FRFS maintains the same superiority in the case of the OneR classifier also. On the other hand, in the case of the C4.5 decision tree, the FSV feature ranking procedure shows better results. However, with just three features, the proposed FRFS scheme shows a similar result as
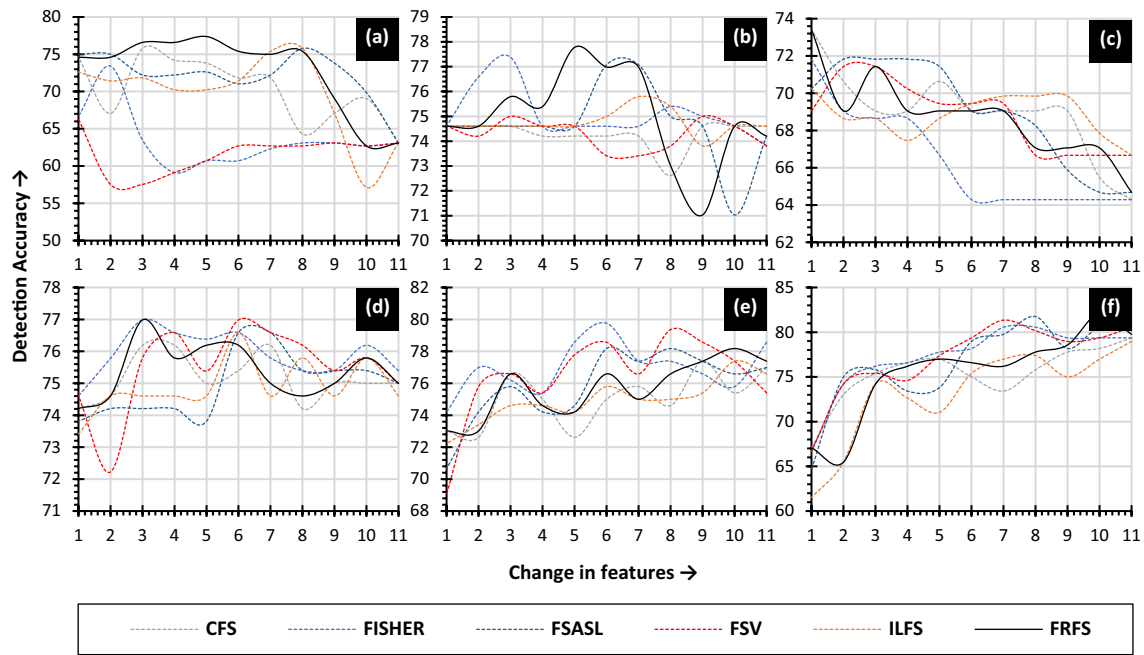
Fig. 4 Classification accuracy of the FRFS-DTN and other peer feature ranking schemes through **a** Naïve Bayes **b** C4.5 **c** OneR **d** SysFor **e** ForestPA **f** Random Forest classifiers on time frequency features of Sakar et al. (2019) dataset

FSV. In a nutshell, FRFS reveals better results while Naïve Bayes is deployed.

When decision forests are explored on the ranked BF segment, it is observed that all the decision forests reveal superior accuracy on a moderate number of BF ($7 < m < 14$). The only variation was observed with the FSV feature ranking scheme under SysFor. In the case of SysFor, the classifier took only few number of features ($1 < m < 21$) in contrast to the FRFS scheme to show the highest detection accuracy. However, as soon as the number of features is added for evaluation, the FRFS supersedes all its peers. It is worth mentioning that the FSV scheme also tried to touch the peak detection rate with the help of other decision forests like ForestPA and Random Forest, but it fails to compete with the proposed FRFS scheme. With just 8 BF in hand, the ForestPA shows the highest accuracy of 78.57%. Conversely, a better detection rate of 79.76% has been observed with Random Forest on nine features, which founds to be slightly more than that of ForestPA.

On VFF ranked features, when Naïve Bayes, C4.5, and OneR were examined, it is observed that the proposed scheme enforced Naïve Bayes to show an impressive detection rate with a small number of features. However, with the increase in features ($m > 5$) all the feature ranking schemes except FSASL shows better result for Naïve Bayes classification. On the contrary, for the C4.5 decision tree and OneR classifier, the proposed FRFS scheme ranked the VFF better so that the concerned classifier gets the peak detection accuracy with few features in hand.

The VFF features on the decision forest show the outstanding result for the FRFS feature ranking scheme. The proposed FRFS scheme works magnificently with the SysFor classifier. The SysFor took a minimal number of FRFS ranked VFF features ($m < 8$) to reach the peak detection rate. Conversely, both Random Forest and ForestPA took the moderate number of features ($8 < m < 15$) to get peak accuracy. One interesting inference observed here is about the proposed FRFS feature ranking scheme. The FRFS works reasonably well with decision forest. A similar inference was also observed in the case of the BF segment.

Figure 4a, b, and c shows the detection accuracy of Naïve Bayes, C4.5 decision tree, and OneR classifiers on the ranked TFF. Like BF and VFF, FRFS ranked TFF, which shows promising detection accuracy than its peer feature ranking schemes. However, on Naïve Bayes, only the CFS and FSASL have the competitive result at par with the FRFS scheme. Similarly, classification on OneR, the detection result is similar except for Fisher, where the Fisher scheme underperforms heavily compared to the proposed FRFS scheme. On the other hand, for the C4.5 decision tree, the FRFS scheme shows better results than all the other feature ranking schemes.

While ranking the features for decision forests classification, a mixed result has been observed. For the SysFor classifier, the proposed FRFS scheme shows better results than CFS, FSASL, and ILFS feature ranking. However, the proposed scheme has not been convincing as compared to FISHER and FSV. Both for ForestPA and Random Forest,

Table 3 Classification accuracy of classifiers on selected features of FRFS on Sakar et al (2019) dataset

Acoustic feature segment	FRFS features selected	Naïve Bayes	C4.5	OneR	SysFor	ForestPA	Random forest
Baseline	3	74.60	72.22	74.21	**77.38**	77.38	74.21
Time frequency	1	74.60	74.60	73.41	**74.21**	73.02	67.06
Vocal fold	6	74.60	71.43	67.46	**76.59**	75.40	74.60

a remarkable result has been observed. The performance output of ForestPA and Random Forest was observed to be linear. Therefore, on these classifiers, feature ranking is not advisable. As the number of TFF is less, considering all features would be a better choice. The feature ranking analysis shows that our proposed FRFS feature ranking scheme shows a promising rank of BF, VFF, and TFF segments. The FRFS scheme has been utilized at the pre-processing stage of the proposed Parkinson Detection System (PDS), where the FRFS ranks the features before the training and detection of the system took place. In this way, the proposed PDS detection capability improved, and the system able to detect incoming PD subjects with utmost precision.

4.2 FRFS feature subset selection

The strength of proposed FRFS scheme is its support for both feature ranking and feature selection. Various feature ranking schemes compared against our proposed FRFS scheme cannot select a subset of features. The feature subset selection procedure for such schemes has been taken care manually by the users. The manual feature subset selection is not always feasible. Our FRFS feature selection scheme undertakes the automatic feature selection using the procedure described in Sect. 3.2.2. In this section, using the FRFS feature subset selection scheme, requisite features of BF, VFF, and TFF segments were selected from Sakar et al. (2019) dataset. Further, those selected features are sent for classifications using the classifiers discussed in Sect. 4.1. The classification output of those classifiers is presented in Table 3.

According to the result obtained in Table 3, the SysFor is emerging as the ideal classifier for all the three acoustic feature groups. The SysFor shows 77.38%, 74.21%, 76.59% detection accuracy for BF, TFF, and VFF. Nevertheless, both Naïve Bayes and C4.5 decision trees also show superior detection accuracy for TFF. On the other hand, the ForestPA shows equivalent detection accuracy of SysFor for the BF segment.

4.3 Parkinson's detection using FRFS on ForEx++

Till this time, a feature ranking and feature selection module has been devised. It has been observed that both the feature ranking and feature selection module of the FRFS scheme efficiently rank the features better and select the most suitable features thereof. In this section, the proposed Parkinson's detection modules have been tested and evaluated, combining the ForEx++ environment on SysFor and ForestPA classifiers separately. In the detection modules, FRFS feature ranking and section modules are attached as a filter module. As a result, the detector works on the outstanding features and detects Parkinson's more efficiently. Ultimately, four Parkinson's detection scheme comes into action; viz., FRFS ranked features with SysFor on ForEx++, FRFS ranked features with ForestPA on ForEx++, FRFS features subset with SysFor on ForEx++, and FRFS features subset with ForestPA on ForEx++. As both the ForestPA and SysFor classifiers work on decision tree ensemble, the number of decision trees in these forests during the detector's training (Pramanik et al. 2021a, b) were controlled to improve detection accuracy. The result is obtained separately through Leave One Subject Out (LOSO) cross-validation, tenfold cross-validation, and an ideal training–testing split. In LOSO, the entire dataset having FRFS ranked features and subset of features is typically divided into k number of blocks. The model is trained on $k-1$ blocks and the kth instance has been used for testing the model. The process continues for $k-1$ times. The same principle is repeated for tenfold cross-validation, where the entire datasets are divided into ten blocks ($k=10$). The proposed PDS performance under LOSO cross-validation has been summarized in Table 4 and Table 5 separately for SysFor and ForestPA classifiers on the ForEx++ framework. Similarly, the performance of the proposed PDS under tenfold cross-validation has been outlined in Table 6 and Table 7 for SysFor and ForestPA classifiers on the ForEx++ framework, respectively.

It should be noted that the theme of the proposed Parkinson's detection model is the decision forests (ForestPA and SysFor), therefore the number of decision trees (t) were added to the forest incrementally, and performance the detection model is observed with change in number of trees. The volume of trees in the forest where the maximum detection accuracy is observed is considered to be the threshold point and the detection model is settled at that point with the desired number of trees.

Table 4 Performance of the proposed Parkinson's Detection System on ForEx++ (SysFor) framework using FRFS ranked features and features subset using acoustic feature groups obtained through LOSO Cross-Validation on Sakar et al. (2019) dataset

ForEx++ (SysFor)	FRFS APPROACH													
	Ranking							Subset						
	m	t	ACC	MAE	RMSE	AUC	PRC	m	t	ACC	MAE	RMSE	AUC	PRC
Baseline	21	1	78.57	0.21	0.46	0.60	0.68	3	2	78.57	0.23	0.44	0.66	0.72
Time frequency	11	2	77.78	0.26	0.47	0.67	0.71	1	2	74.60	0.26	0.51	0.49	0.62
Vocal fold	22	2	78.97	0.24	0.46	0.62	0.69	6	56	75.00	0.26	0.46	0.67	0.72
Combined	54	62	82.14	0.23	0.39	0.77	0.81	7	2	79.37	0.24	0.45	0.63	0.71

Table 5 Performance of the proposed Parkinson's Detection System on ForEx++ (ForestPA) framework using FRFS ranked features and features subset through LOSO Cross-Validation on Sakar et al. (2019) dataset

ForEx++ (ForestPA)	FRFS Approach													
	Ranking							Subset						
	m	t	ACC	MAE	RMSE	AUC	PRC	m	t	ACC	MAE	RMSE	AUC	PRC
Baseline	21	12	78.97	0.31	0.40	0.71	0.77	3	11	76.59	0.33	0.41	0.71	0.76
Time frequency	11	33	78.17	0.33	0.40	0.72	0.77	1	9	74.21	0.36	0.44	0.55	0.67
Vocal fold	22	8	80.56	0.32	0.40	0.71	0.77	6	9	76.59	0.35	0.43	0.61	0.70
Combined	54	10	80.16	0.31	0.40	0.73	0.79	7	8	76.59	0.34	0.42	0.64	0.72

Table 6 Performance of the proposed Parkinson's detection system on ForEx++ (SysFor) framework using FRFS ranked features and features subset through 10-Fold Cross-Validation on Sakar et al (2019) dataset

ForEx++ (SysFor)	FRFS approach													
	Ranking							Subset						
	m	t	ACC	MAE	RMSE	AUC	PRC	m	t	ACC	MAE	RMSE	AUC	PRC
Baseline	21	19	76.19	0.25	0.44	0.7	0.74	3	19	77.78	0.26	0.44	0.71	0.75
Time frequency	11	4	77.78	0.24	0.43	0.71	0.74	1	4	74.60	0.26	0.50	0.51	0.63
Vocal fold	22	12	80.56	0.23	0.42	0.69	0.75	6	12	77.38	0.25	0.45	0.65	0.71
Combined	54	94	83.73	0.23	0.39	0.78	0.82	7	94	78.17	0.24	0.46	0.63	0.70

Table 7 Performance of the proposed Parkinson's detection system on ForEx++ (ForestPA) framework using FRFS ranked features and features subset through 10-fold cross-validation on Sakar et al. (2019) dataset

ForEx++ (ForestPA)	FRFS approach													
	Ranking							Subset						
	m	t	ACC	MAE	RMSE	AUC	PRC	m	t	ACC	MAE	RMSE	AUC	PRC
Baseline	21	30	78.57	0.31	0.40	0.76	0.79	3	30	79.76	0.31	0.39	0.74	0.78
Time frequency	11	13	78.57	0.33	0.41	0.72	0.77	1	13	75.00	0.36	0.44	0.55	0.66
Vocal fold	22	22	80.16	0.32	0.40	0.7	0.75	6	22	76.59	0.35	0.42	0.63	0.71
Combined	54	32	79.76	0.31	0.40	0.71	0.78	7	32	78.17	0.34	0.41	0.72	0.78

4.3.1 Leave one subject out (LOSO) cross-validation

Table 4 represents the LOSO cross-validation of the proposed PD model under the SysFor classifier. When the Sys-For classifier is employed, the model detects the presence of Parkinson's with a detection accuracy of 78.57%, 77.78%, 78.97% for BF, TFF, and VFF when trained on FRFS ranked features. Considering the BF, the FRFS ranked features, and the FRFS features subset reveals equal detection accuracy. The former results in a low mean absolute error, whereas

the latter produced a superior AUC. The improved AUC and the convincing detection accuracy of 78.57% reveal only three BF to detect Parkinson's with better time complexity. For TFF and VFF, the FRFS ranking is the better choice, yielding 3.18% and 3.97% more detection accuracy than the FRFS feature subset. When all the acoustic feature groups are combined and ranked as per FRFS, the model is effective with an 82.14% detection rate. However, to achieve such convincing detection accuracy, the SysFor decision forest under ForEx++ frameworks took the help of 62 decision trees.

On the other hand, when the ForestPA has been plugged into the ForEx++ environment for Parkinson's Detection, the FRFS ranked features demonstrate improved results on the BF, TFF, and VFF segments (Table 5) compared to earlier SysFor. However, to achieve better detection accuracy, the ForestPA undertakes a significant number of decision trees. However, combining all the features and ranking through FRFS, the detector took only 10 decision trees with a detection accuracy of 80.16%. In ForestPA, the FRFS feature ranking proved to be a better choice than the FRFS feature subset selection. In feature subset selection, the BF are alone the better choice than the TFF, VFF, and even combined features. The FRFS feature subset helps the detector score the highest detection accuracy of 76.59%, with the lowest error rates.

In LOSO cross-validation, both the SysFor and ForestPA decision forests show promising detection results on all feature groups. Both the detectors work smartly on FRFS ranked features and FRFS feature subset. At the next stage of analysis, a tenfold cross-validation test was also conducted to confirm the performance of the proposed approach.

4.3.2 Ten-fold cross-validation

Under tenfold cross-validation, the results obtained for SysFor and ForestPA classifiers on the ForEx++ framework have been presented in Tables 6 and 7. Observing Table 6 for SysFor decision forest, the performance received for acoustic features is convincing, but it is not par with the performance received during the LOSO cross-validation. The detection accuracy received in tenfold cross-validation through FRFS ranked BF, and VFF features are 2.38% and 1.59% lower than the LOSO cross-validation. The performance of SysFor remains unchanged for TFF. On the other hand, the SysFor shows improved detection accuracy with just the FRFS feature subset of the VFF.

The result is flipped, combining all the feature groups for conducting tenfold cross-validation on SysFor decision forest. The tenfold validation revealed improved detection accuracy of 83.73% for FRFS ranked features in contrast to LOSO cross-validation. On the contrary, the detection accuracy falls when FRFS feature subsets are considered

for tenfold validation. The tenfold cross-validation on FRFS feature subsets shows 78.17% detection accuracy, which is 1.2% lower than the LOSO cross-validation.

A similar result has been produced for tenfold cross-validation on the tenfold cross-validation of ForestPA classifier on ForEx++ framework. The result obtained is a bit opposite to the SysFor. The detection accuracy achieved for FRFS ranked features of the combined feature segment shows a lower detection rate than the similar LOSO cross-validation. Not only that, when a subset of combined features is selected through FRFS, the detection accuracy is recorded more than the LOSO validation. A similar kind of mixed result has been ascertained for all the feature groups.

Both the LOSO and tenfold cross-validation revealed mixed results and mostly opposite to each other. Therefore, a third level of the test has been conducted using a training–testing split. However, the training–testing split process has been automated and is not left to the users. Iteratively, the training instances are increased, and the testing instances are decreased. This process has been continued till a breakeven point where the highest detection result has been received.

4.3.3 Validation through training–testing instance split

The training–testing split of instances on different acoustic feature groups and combined feature groups are sent for evaluation both for SysFor and ForestPA, which yields four distinct PD detection systems. The proposed Parkinson's detections are named Method I (ForEx++ based SysFor detection module on FRFS ranked features), Method II (ForEx++ based SysFor detection module on FRFS selected features), Method III (ForEx++ based ForestPA detection module on FRFS ranked features), and Method IV (ForEx++ based ForestPA detection module on FRFS selected features). The performance outcome of all four types of PDS is presented in Tables 8, 9, 10, 11.

For the Method I Parkinson's detection, the SysFor classifier within the ForEx++ environment has been presented with the settings and results presented in Table 8. According to Table 8, the BF demonstrate the highest detection accuracy of 92.45%, claiming the highest AUC of 0.91. The training–testing split of 79%-21% is ideal with 32 decision tree ensembles to achieve such an impressive accuracy rate. However, combining all the feature groups, the decision trees of SysFor has been consolidated to only 3. Therefore, though the detection accuracy is slightly lower than that of the detection accuracy achieved by FRFS ranked BF, combining all the features reduced the false positives significantly. But looking towards the detection accuracy and better AUC, the Method I Parkinson's detection system has been proposed with FRFS ranked BF having 32 decision trees in the SysFor forest.

Table 8 Performance of the proposed ForEx++ (SysFor)+FRFS (ranked features) Parkinson's detection system through an ideal training–testing instance split (Method I) on Sakar et al (2019) dataset

ForEx++ (SysFor)	m	Tr (%)	Ts (%)	t	ACC	MAE	RMSE	FNR	FPR	SPE	SEN	AUC	PRC
Baseline	21	79	21	32	92.45	0.14	0.28	0.08	0.25	0.75	0.92	0.91	0.94
Time Frequency	11	69	31	4	84.81	0.23	0.40	0.15	0.55	0.45	0.85	0.66	0.78
Vocal Fold	22	79	21	2	88.68	0.16	0.32	0.11	0.41	0.59	0.89	0.84	0.88
Combined	54	79	21	3	90.57	0.16	0.29	0.09	0.18	0.82	0.91	0.90	0.92

Table 9 Performance of the proposed ForEx++ (SysFor)+FRFS (features subset) Parkinson's detection system through an ideal training–testing instance split (Method II) on Sakar et al. (2019) dataset

ForEx++ (SysFor)	m	Tr (%)	Ts (%)	t	ACC	MAE	RMSE	FNR	FPR	SPE	SEN	AUC	PRC
Baseline	3	78	22	28	89.29	0.17	0.33	0.11	0.37	0.63	0.89	0.88	0.91
Time frequency	1	79	21	1	81.13	0.19	0.43	0.19	0.81	0.19	0.81	0.50	0.69
Vocal fold	6	79	21	84	84.91	0.18	0.36	0.15	0.50	0.50	0.85	0.73	0.82
Combined	7	79	21	3	86.79	0.18	0.34	0.13	0.49	0.51	0.87	0.79	0.84

Table 10 Performance of the proposed ForEx++ (ForestPA)+FRFS (ranked features) Parkinson's detection system through an ideal training–testing instance split (Method III) on Sakar et al. (2019) dataset

ForEx++ (ForestPA)	m	Tr (%)	Ts (%)	t	ACC	MAE	RMSE	FNR	FPR	SPE	SEN	AUC	PRC
Baseline	21	80	20	36	94.12	0.25	0.29	0.06	0.24	0.76	0.94	0.97	0.98
Time frequency	11	78	22	90	87.50	0.33	0.38	0.13	0.51	0.49	0.88	0.68	0.82
Vocal fold	22	80	20	9	92.16	0.26	0.32	0.08	0.32	0.68	0.92	0.87	0.92
Combined	54	69	31	16	91.14	0.29	0.35	0.09	0.32	0.68	0.91	0.73	0.82

Table 11 Performance of the proposed ForEx++ (ForestPA)+FRFS (features subset) Parkinson's detection system through an ideal training–testing instance split (Method IV) on Sakar et al. (2019) dataset

ForEx++ (ForestPA)	m	Tr (%)	Ts (%)	t	ACC	MAE	RMSE	FNR	FPR	SPE	SEN	AUC	PRC
Baseline	3	80	20	2	90.20	0.22	0.31	0.10	0.33	0.67	0.90	0.85	0.89
Time frequency	1	79	21	6	83.02	0.34	0.38	0.17	0.65	0.35	0.83	0.71	0.80
Vocal fold	6	78	22	81	87.50	0.32	0.37	0.13	0.51	0.49	0.88	0.68	0.81
Combined	7	79	21	22	92.45	0.29	0.34	0.08	0.32	0.68	0.92	0.84	0.91

Method II Parkinson's detection system has been proposed using the same ForEx++ based SysFor decision forest on the FRFS feature subset. The proposed approach undertakes only 3, 1, and 6 features of BF, TFF, and VFF features. Like Method I, the Method II detection approach favors BF features to reveal 89.29% of detection accuracy with good sensitivity. However, all other feature groups and even combining all feature groups are not convincing for Method II Parkinson detection. Therefore, the Method II Parkinson detection is also proposed with FRFS ranked BF features subset with 28 decision trees in the ForEx++ based SysFor decision forest.

The Method III Parkinson's detection system has been identified with FRFS ranked BF features at the third stage. The detector used in this case is the ForestPA in the ForEx++ environment. The ForEx++ based ForestPA decision forest reveals the highest ever 94.12% detection accuracy with the best sensitivity rate of 0.94. The AUC is also 0.97, which is also promising. The Method II approach takes the help of 36 decision trees to achieve such a magnificent detection rate. Combining all the feature segments for ForEx++ based SysFor, the detection rate achieved (91.14%) is better than what was achieved (90.57%) while employing the same ranked combine features on ForEx++
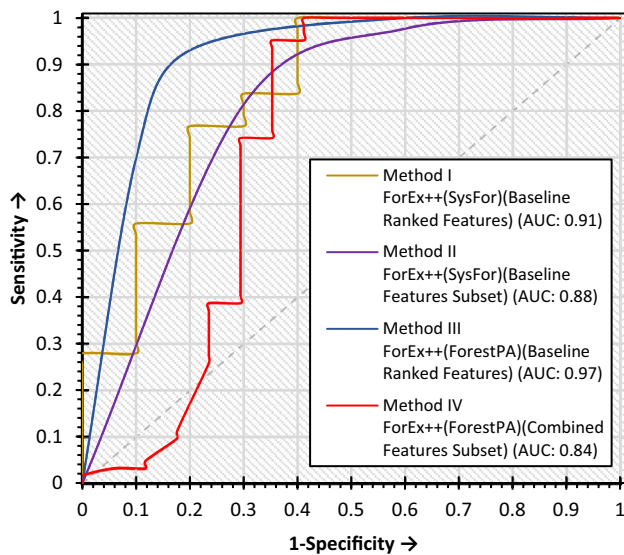
Fig. 5 Receiver Operating Curve (ROC) analysis of the proposed methods on Sakar et al. (2019) dataset

based SysFor detector. In a nutshell, the Method III Parkinson detection has been proposed with FRFS ranked BF features on ForEx++ based ForestPA decision forest having 36 trees in the forest.

Finally, Method IV Parkinson's detection approach has been devised on ForEx++ based ForestPA decision forest, combining and selecting only seven features through FRFS. It is because individual feature groups are not promising as compared to the combined result. However, the results were obtained for individual feature groups than the ForEx++ based SysFor approach.

Observing the performance of the four methods outlined in Tables 8, 9, 10, 11, it is relatively easy to conclude that the detection accuracy of the proposed methods lies in the order Method III, Method IV, Method I, and Method II, Where Method III is the top performer with highest detection accuracy. Method II is the most diminutive performer with the lowest detection accuracy. Both Method IV and Method I are equal contenders. Still, priority has been given to Method IV over Method I because Method IV outperforms with just seven hand features. However, Method I have the lowest mean absolute error than Method IV. Therefore, a Receiver Operating Curve (ROC) analysis has been conducted to get a precise result. A ROC analysis proved to be a great aid in medical data classification (Hajian-Tilaki 2013). The AUC represents the classification's intrinsic ability to discriminate between the diseased and healthy subjects (Metz 1978). The ROC curve and corresponding AUC have been presented in Fig. 5 for all four methods.

Figure 5 shows that the ROC of Method III is highest and occupies a maximum ROC of 0.97, and with the AUC of 0.84, the ROC lies at the lowest for Method IV.

The ROC of Method I and Method II were recorded as 0.91 and 0.88, respectively. Therefore, based on the AUC obtained, the Method I and Method III found excellent (AUC ≥ 0.91) and Method II and Method IV is a good choice (0.8 ≤ AUC ≥ 0.9) (Srivastava 2019).

At the next stage of results, the detection output has been visualized through concentration graphs. It allows us to visualize how effective the proposed models. According to the concentration analysis, a model is said to be effective if it concentrates more towards True Positives (TP) and True Negatives (TN) than False Positives (FP) and False Negatives (FN). The concentration graphs for all four models have been presented in Fig. 6. According to Fig. 6, the dark green and light green areas' concentration looks more for all the proposed methods. Especially for Method III, the concentration of false positive is the lowest, proving Method III is the best approach for Parkinson's detection. The Penalizing Decision Forest (ForestPA) in a ForEx++ framework with FRFS ranked BF segment seems most effective for Parkinson's detection with just 36 decision trees in the forest.

4.3.4 Comparing the proposed detection model with existing recent methods

The detailed validation has been conducted so far on the proposed Parkinson's detection methods. The proposed methods stand tall in all kinds of validation approaches. An effort has been made to compare our methods with existing state-of-the-art models. Since all the proposed methods are based on the supervised classification principle, the models shortlisted from the literature reviewed are also based on supervised classification techniques. An unbiased comparison has been conducted by implementing the proposed approaches separately both in the unbalanced Sakar et al. (2013) and Sakar et al. (2019) datasets and the balanced (Naranjo et al. 2016) dataset. Implementing the proposed models in multiple datasets have two benefits, (a) It provides a scope to understand how the proposed models are behaving both in a class balanced and unbalanced environment (b) The strengths and weaknesses of the proposed models will be well figured out by comparing with the existing approaches on a specific to datasets. The existing Parkinson's detection approaches and their detection outcomes along with the outcomes of the proposed methods, have been presented in Tables 12, 13, and 14. The comparison has been conducted in terms of detection accuracy, sensitivity, and specificity.

All the approaches of Table 12 have been implemented on the Sakar et al. (2013) dataset; thus, providing an unbiased environment for comparison methods. The proposed approach FRFS (Subset) + ForEx++ (SysFor) (t = 40) (Method II) has been evolved as the detector having the highest detection accuracy and sensitivity of 92.86%. The sensitivity of the proposed Method II reveals that
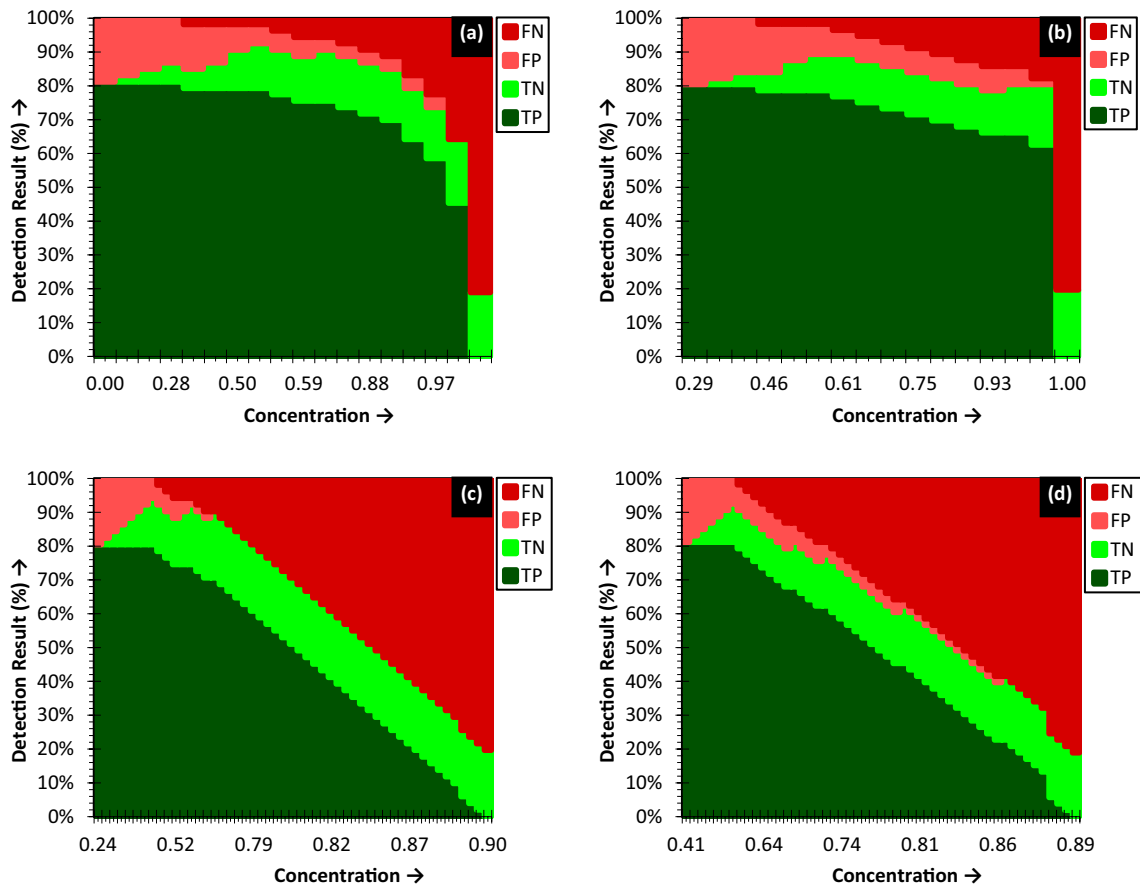
Fig. 6 Detection result with the concentration of the proposed methods **a** ForEx++ (SysFor)+FRFS (ranked features) (Method I) (*t=32, m=21*) on BF segment **b** ForEx++ (SysFor)+FRFS (features subset) (Method II) (*t=28, m=3*) on BF segment **c** ForEx++ (ForestPA)+FRFS (ranked features) (Method III) (*t=36, m=21*) on BF segment **d** ForEx++ (ForestPA)+FRFS (features subset) (Method IV) (*t=22, m=7*) on combined features on Sakar et al. (2019) dataset

Table 12 Performance comparison of the proposed Parkinson's detection methods along with existing approaches on Sakar et al. (2013) dataset

Existing and proposed methods	ACC (%)	SEN (%)	SPE (%)
Ensemble of k-NN, NB and SVM + A-MCFS (Behroozi and Sami 2016)	82.50	80.00	85.00
k-NN with GA + LDA (Ali et al. 2019a, b)	92.50	95.00	90.00
ELM (Log Sigmoid) (Agarwal et al. 2016)	87.50	85.71	89.47
Ensemble of RF, SVM, ELM (Li et al. 2017)	87.50	80.00	95.00
Multiple neural network ensemble (Berus et al. 2018)	81.33	76.33	86.33
MENN + RF (Zhang et al. 2016)	81.50	70.50	92.50
k-Nearest Neighbour (k=3) (Cantürk and Karabiber 2016)	68.75	70.60	66.90
FRFS (Ranked) + ForEx++ (SysFor) (t=32) (Method I)	88.24	88.24	45.10
FRFS (Subset) + ForEx++ (SysFor) (t=40) (Method II)	92.86	92.86	57.14
FRFS (Ranked) + ForEx++ (ForestPA) (t=9) (Method III)	87.50	87.50	71.47
FRFS (Subset) + ForEx++ (ForestPA) (t=14) (Method IV)	88.24	88.24	45.10

92.86% of all Parkinson's subjects are detected successfully, and 7.14% of positive subjects are passed the test as controls. On the other hand, the proposed Method II is suffering for control subjects on the ground of specificity.

The specificity shows that 57.14% of the control subjects are correctly identified as controls, whereas 42.86% of the control subjects are detected as Parkinson's. However, Method III detects the control subjects with satisfied

Table 13 Performance comparison of the proposed Parkinson's detection methods along with existing approaches on Sakar et al. (2019) dataset

Existing and proposed methods	ACC (%)	SEN (%)	SPE (%)
OAA + SVM (Polat and Nour 2020)	83.75	81.67	85.83
OAA + K-NN (Polat and Nour 2020)	82.08	79.17	85.00
OAA + LR (Polat and Nour 2020)	77.50	80.00	75.00
RBF (Upadhya et al. 2019)	83.80	97.50	70.00
SAE (LDA + all features) (Xiong and Lu 2020)	76.00	84.00	89.00
SAE (LDA + CFS) (Xiong and Lu 2020)	89.00	92.00	93.00
SAE (RF + RFE) (Xiong and Lu 2020)	81.00	84.00	86.00
SAE (LDA + MRMR) (Xiong and Lu 2020)	91.00	92.00	94.00
Support vector machine (Bourouhou Et Al. 2016)	80.00	77.00	83.00
Random forest (Smekal et al. 2015)	71.43	72.60	69.40
Classification and regression trees (Mekyska et al. 2015)	75.19	79.80	67.40
Gaussian mixtures models (Jeancolas et al. 2017)	48.50	55.30	39.30
FRFS (RANKED) + FOREX++ (SYSFOR) (T = 32) (Method I)	92.45	92.00	75.00
FRFS (SUBSET) + FOREX++ (SYSFOR) (T = 28) (Method II)	89.29	89.00	63.00
FRFS (RANKED) + FOREX++ (FORESTPA) (T = 36) (Method III)	94.12	94.00	76.00
FRFS (SUBSET) + FOREX++ (FORESTPA) (T = 22) (Method IV)	92.45	92.00	68.00

Table 14 Performance comparison of the proposed Parkinson's detection methods along with existing approaches on Naranjo et al. (2016) dataset

Methods implemented on Naranzo et al. (2016) dataset***	ACC (%)	SEN (%)	SPE (%)
CPC (Group) + FMM (Naranjo et al. 2019)	83.60	82.70	84.70
CPC (Joint) + FMM (Naranjo et al. 2019)	82.20	82.10	83.70
Ensemble k-NN + ReliefF (Yaman et al. 2020)	90.42	90.83	90.00
Cosine k-NN + ReliefF (Yaman et al. 2020)	91.25	93.33	89.17
Gaussian SVM + ReliefF (Yaman et al. 2020)	91.25	90.83	91.67
Quadratic SVM + ReliefF (Yaman et al. 2020)	87.50	86.67	88.33
Radial basis function (Perez et al. 2016)	85.25	90.23	80.28
Fuzzy neural network (Guimarães et al. 2020)	80.88	85.34	78.75
Radial basis function (Viswanathan et al. 2018)	70.00	67.00	73.00
FRFS (Ranked) + ForEx++ (SysFor) (t = 25) (Method I)	96.88	96.88	96.88
FRFS (Subset) + ForEx++ (SysFor) (t = 99) (Method II)	95.65	95.65	96.01
FRFS (Ranked) + ForEx++ (ForestPA) (t = 11) (Method III)	96.67	96.67	96.67
FRFS (Subset) + ForEx++ (ForestPA) (t = 14) (Method IV)	96.67	96.67	96.67

specificity of 71.47%. Among the other approaches, the ensemble of RF, SVM, ELM (Li et al. 2017) detects the controls as control with the highest ever 95.00%. It can be seen from Table 13 that the k-NN with GA + LDA (Ali et al. 2019a, b) shows the best result of detection among all other approaches, including our proposed approaches. Nevertheless, proposed approaches show similar or better results than the peer detection models.

The proposed approaches are tested with many other existing Parkinson's detection models at the next level of comparison on the Sakar et al. (2019)) dataset. It is interesting to notice the performance of the proposed model in a recent and modern acoustic dataset having more instances than Sakar et al. (2013) dataset. The detection results of our proposed models with other peer models on the Sakar et al. (2019) dataset have been presented in Table 13.

On the Sakar et al. (2019) dataset, the proposed FRFS (Ranked) + ForEx++ (ForestPA) (t = 36) (Method III) model shows the highest detection accuracy of 94.12%. The ForestPA classifier took only 36 decision trees to gather sensitivity and specificity of 94.00% and 76.00%. The FRFS ranked features proved to be a better choice in Sakar et al.'s (2019) dataset. The SAE (LDA + mRMR) (Xiong and Lu 2020) approach also shows consistent results in terms of detection accuracy, specificity, and sensitivity. It should be noted that the proposed approaches are still showing better results than other peer models.

In the final level of comparative analysis, the proposed model again tested with few more existing models on Naranjo et al. (2016) dataset. It is worth watching the models' performance because the Naranjo et al. (2016) dataset is a balanced dataset having an equal number of Parkinson's

and controls subjects. The result of the proposed models with other peer models has been presented in Table 14.

Table 14 shows, the FRFS ranked features are a better choice than the FRFS feature subsets. The Method I of the proposed work shows the highest detection accuracy, sensitivity, and specificity of 96.88%. In a landscape, the performance of Method I, Method II, Method III, and Method IV appears to be far ahead of other recent models implemented on Naranjo et al. (2016) dataset.

The proposed four methods of Parkinson's' detection are implemented on three acoustic datasets (Sakar et al. 2013, 2019; Naranjo et al. 2016). Many more recent Parkinson's detection approaches have been compared based on detection accuracy, sensitivity, and specificity. In all the datasets, the proposed methods come up with excellent results. Nevertheless, few observations or inferences have been evolved. From the comparisons, it is evident that the proposed approaches respond to the control subjects in the case of Naranjo et al. (2016) dataset, where the specificity gears up to 96.88%. On the other hand, Method I suffer miserably for Sakar et al. (2013) dataset with the lowest ever specificity. The reason for low specificity is the inability of SysFor to score more in a smaller training sample. It should be noted that the Sakar et al. (2013) dataset holds only 68 instances having 48 Parkinson's and 20 controls. Due to the low number of instances with a high-class imbalance ratio, the SysFor becomes biased towards the Parkinson's subjects, resulting in lower specificity. A similar type of results is observed for Sakar et al. (2019) dataset. Although the dataset holds acoustic information of the adequate number of subjects, but the class imbalance issues subside the performance of the proposed PDSs. But this was not the case with Naranjo et al. (2016) dataset. The Naranjo et al. (2016) dataset holds acoustic information of 80 subjects comprising of 40 controls and 40 Parkinson's. The perfect class imbalance ratio of Naranjo et al. (2016) dataset shows an exemplary performance of proposed Method I. The Method I show highest detection accuracy, sensitivity and specificity of 96.88%.

5 Conclusion

This article presented four distinct approaches to Parkinson's Disease detection methods. The methods rely on acoustic voice signal data to detect the presence of Parkinson's in its early stage. The proposed methods have been developed on prominent acoustic features. The acoustic feature groups are ranked through a new feature selection scheme known as Feature Ranking to Feature Selection (FRFS) via Directed Tests of Normality (FRFS-DTN). The FRFS scheme ranked the features based on the normality score and selects only those normally distributed features and contributed most towards classification. The four Parkinson's Detection

Methods proposed here have been developed through SysFor and ForestPA decision forest algorithms through a state-of-the-art ForEx++ framework. The ForEx++ framework has been employed to improve the decision forest building process, thus improving detection. The volume of trees in the decision forests of SysFor and ForestPA is controlled during the training process, which yields significant detection accuracy. All the four proposed methods are validated separately through tenfold cross-validation, LOSO cross-validation, and through an ideal training–testing split of instances. All four methods are proposed with the minimum number of decision trees in the forest. Method III found most effective through ForestPA decision forest having 36 decision trees when trained and tested on FRFS ranked BF segment. In the end, a comparative analysis has also been conducted with recent Parkinson detection approaches, where all the four methods found better efficiency in segregating Parkinson's subjects from controls.

Like any other PDS, the proposed methods also have limitations; if addressed, detection efficiency can be improved further. The proposed methods do not consider a feedback approach to the training module. A feedback mechanism to the training module essentially brings more dynamism to the proposed methods. The proposed methods can be modelled to detect the stages of Parkinson's, which helps to identify the severity of the disease. Moreover, the methods can be extended following Unified Parkinson's Disease Rating Scale to map the disease severity level. The baseline, time frequency, and vocal fold features are considered for Parkinson's detection in this proposed work. However, the MFCC, wavelet and TQWT features contain more informative vocal features. Therefore, modern function-based ensemble approaches can be explored on these feature groups for better detection results. The age and gender of the subjects are not considered in the article. However, vocal features like pitch, voice intensity and detrended fluctuation have a different curve for different gender and age groups. Therefore, the proposed model can be improvised further to detect Parkinson's in different age groups and genders. The shorter length of phonation in PD could be another factor influencing voice analysis. In future work, the proposed model can be tested on phonation of varying duration.

Code availability (software application or custom code) Not applicable.

Declarations

Conflict of interest None.

Ethicl approval Not applicable' as the study did not require ethical approval.

Consent to participate Not applicable' as the studies not involving humans.

Consent for publication As per the journal guidelines and norms.

References

Afonso LCS, Rosa GH, Pereira CR, Weber SAT, Hook C, AlbuquerquePapa VCHJP (2019) A recurrence plot-based approach for Parkinson's disease identification. Futur Gener Comput Syst 94:282–292

Agarwal A, Chandrayan S, Sahu SS (2016) Prediction of Parkinson's disease using speech signal with extreme learning machine. Int Conf Electr Electron Optim Techn (ICEEOT). https://doi.org/10.1109/ICEEOT.2016.7755419

Aich S, Younga K, Hui KL, Al-Absi AA, Sain M (2018) A nonlinear decision tree based classification approach to predict the Parkinson's disease using different feature sets of voice data." In: 2018 20th International Conference on Advanced Communication Technology (ICACT), pp 638–42

Ali L, Zhu Ce, Zhang Z, Liu Y (2019a) Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network. IEEE J Transl Eng Health Med 7:1–10. https://doi.org/10.1109/JTEHM.2019.2940900

Ali L, Zhu Ce, Zhou M, Liu Y (2019b) Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection. Expert Syst Appl 137:22–28

Almeida JS, Rebouças PP, Filho TC, Wei W, Damaševičius R, Rytis M, de Albuquerque VHC (2019) Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques. Pattern Recognit Lett 125:55–62

Anand A, Haque MA, Alex JSR, Venkatesan N (2018) Evaluation of machine learning and deep learning algorithms combined with dimentionality reduction techniques for classification of Parkinson's disease. In: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp 342–47

Behroozi M, Sami A (2016) A multiple-classifier framework for Parkinson's disease detection based on various vocal tests. Int J Telemed Appl 2016:1–9. https://doi.org/10.1155/2016/6837498

Benba A, Jilbab A, Hammouch A (2016) Voice analysis for detecting patients with Parkinson's disease using the hybridization of the best acoustic features. Int J Electr Eng Inf 8(1):108–116. https://doi.org/10.15676/ijeei.2016.8.1.8

Berus L, Klancnik S, Brezocnik M, Ficko M (2018) Classifying Parkinson's disease based on acoustic measures using artificial neural networks. Sensors 19(1):16. https://doi.org/10.3390/s19010016

Bhat S, Rajendra Acharya U, Hagiwara Y, Dadmehr N, Adeli H (2018) Parkinson's disease: cause factors, measurable indicators, and early diagnosis. Comput Biol Med 102:234–241

Bonett DG, Seier E (2002) A test of normality with high uniform power. Comput Stat Data Anal 40(3):435–445

Bourouhou A, Jilbab A, Nacir C, Hammouch A (2016) Comparison of classification methods to detect the Parkinson disease. In: 2016 International Conference on Electrical and Information Technologies (ICEIT), pp 421–24

Bradley PS, Mangasarian OL (1998) Feature selection via concave minimization and support vector machines. ICML 98:82–90

Braga D, Madureira AM, Coelho L, Ajith R (2019) Automatic detection of Parkinson's disease based on acoustic analysis of speech. Eng Appl Artif Intell 77:148–158. https://doi.org/10.1016/j.engappai.2018.09.018

Cantürk İ, Karabiber F (2016) A machine learning system for the diagnosis of Parkinson's disease from speech signals and its application to multiple speech signal types. Arab J Sci Eng 41(12):5049–5059. https://doi.org/10.1007/s13369-016-2206-3

Chandrayan S, Agarwal A, Arif M, Sahu SS (2017) Selection of dominant voice features for accurate detection of Parkinson's disease. In: 2017 Third International Conference on Biosignals, Images and Instrumentation (ICBSII), pp 1–4

Cigdem O, Beheshti I, Demirel H (2018) Effects of different covariates and contrasts on classification of Parkinson's disease using structural MRI. Comput Biol Med 99:173–181

D'Agostino RB (1986) Goodness-of-Fit-Techniques, 68th edn. CRC Press, London

Dinesh A, He J (2017) Using machine learning to diagnose Parkinson's disease from voice recordings. In: 2017 IEEE MIT Undergraduate Research Technology Conference (URTC), pp 1–4

Du L, Shen Y-D (2015) Unsupervised feature selection with adaptive structure learning. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 209–18

Edwards S (2008). In: Thomas MC, Joy A (eds) Thomas elements of information theory. John Wiley & Sons, Inc, New Jersy

Ertuğrul ÖF, Yilmaz Y, Tekin R, Almali MN (2016) Detection of Parkinson's disease by shifted one dimensional local binary patterns from gait. Expert Syst Appl 56:156–163. https://doi.org/10.1016/j.eswa.2016.03.018

Farrús M, Hernando J (2009) Using jitter and shimmer in speaker verification. IET Signal Proc 3(4):247. https://doi.org/10.1049/iet-spr.2008.0147

Farrús M, Hernando J, Ejarque P (2007) Jitter and shimmer measurements for speaker recognition. Proc Annu Conf Int Speech Commun Assoc 2:1153–1156

Najmeh F, Samadiani N (2017) Parkinson's disease detection using ensemble techniques and genetic algorithm. In: 2017 Artificial Intelligence and Signal Processing Conference (AISP), pp 162–65

Gel YR, Miao W, Gastwirth JL (2007) Robust directed tests of normality against heavy-tailed alternatives. Comput Stat Data Anal 51(5):2734–2746. https://doi.org/10.1016/j.csda.2006.08.022

Gómez-Vilda P, Mekyska J, Ferrández JM, Palacios-Alonso D, Gómez-Rodellar A, Rodellar-Biarge V, Galaz Z et al (2017) Parkinson

disease detection from speech articulation neuromechanics. Front Neuroinform 11:56

Gottapu RD, Dagli CH (2018) Analysis of Parkinson's disease data. Proced Comput Sci 140:334–341

Gu Q, Li Z, Han J (2012) Generalized fisher score for feature selection. arxiv Preprint ArXiv:1202.3725

Guimarães AJ, de Campos Souza PV, Lughofer E (2020) Hybrid model for Parkinson's Disease Prediction. Springer, Cham, pp 621–634. https://doi.org/10.1007/978-3-030-50143-3_49

Gupta D, Julka A, Jain S, Aggarwal T, AshishArunkumar KN, de Albuquerque VH (2018) Optimized cuttlefish algorithm for diagnosis of Parkinson's disease. Cognit Syst Res 52:36–48

Hajian-Tilaki K (2013) Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Casp J Intern Med 4(2):627–635

Harel B, Cannizzaro M, Snyder PJ (2004) Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. Brain Cognit 56(1):24–29

Islam Z, Giggins H (2011) Knowledge discovery through sysfor: a systematically developed forest of multiple decision trees KDMI: a novel method for missing values imputation using two levels of horizontal partitioning in a data set view project a hybrid clustering technique combining a novel genetic algorithm with k-means view project knowledge discovery through sysfor-a systematically developed forest of multiple decision trees. Researchgate.Net.*** https://www.researchgate.net/publication/236894348

Jeancolas L, Benali H, Benkelfat B-E, Mangone G, Corvol J-C, Vidailhet M, Lehericy S, Petrovska-Delacrétaz D (2017) Automatic detection of early stages of Parkinson's disease through acoustic voice analysis with mel-frequency cepstral coefficients. In: 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pp 1–6

Joshi D, Khajuria A, Joshi P (2017) An automatic non-invasive method for Parkinson's disease classification. Comput Methods Programs Biomed 145:135–145

Karan B, Sahu SS, Mahto K (2020) Parkinson disease prediction using intrinsic mode function based features from speech signal. Biocybern Biomed Eng 40(1):249–264

Kumar R, Indrayan A (2011) Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatr 48(4):277–287

Li Y, Yang L, Wang P, Zhang C, Xiao J, Zhang Y, Qiu M (2017) Classification of Parkinson's disease by decision tree based instance selection and ensemble learning algorithms. J Med Imaging Health Inf 7(2):444–452. https://doi.org/10.1166/jmihi.2017.2033

Little M, McSharry P, Hunter E, Spielman J, Ramig L (2008) Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Nat Preced. https://doi.org/10.1038/npre.2008.2298.1

Mekyska J, Galaz Z, Mzourek Z, Smekal Z, Rektorova I, Eliasova I, Kostalova M et al (2015) Assessing progress of Parkinson's disease using acoustic analysis of phonation. In: 2015 4th International Work Conference on Bioinspired Intelligence (IWOBI), pp 111–18

Metz CE (1978) Basic Principles of ROC analysis. Semin Nucl Med 8(4):283–298. https://doi.org/10.1016/S0001-2998(78)80014-2

Mongia PK, Sharma RK (2014) Estimation and statistical analysis of human voice parameters to investigate the influence of psychological stress and to determine the vocal tract transfer function of an individual. J Comput Netw Commun. https://doi.org/10.1155/2014/290147

Montaña D, Campos-Roca Y, Pérez CJ (2018) A Diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease. Comput Methods Programs Biomed 154:89–97

Mostafa SA, Mustapha A, Mohammed MA, Hamed RI, Arunkumar N, Ghani MKA, Jaber MM, Khaleefah SH (2019) Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease. Cogn Syst Res 54:90–99

Naranjo L, Pérez CJ, Campos-Roca Y, Martín J (2016) Addressing voice recording replications for Parkinson's disease detection. Expert Syst Appl 46:286–292. https://doi.org/10.1016/j.eswa.2015.10.034

Naranjo L, Pérez CJ, Martín J, Campos-Roca Y (2017) A Two-Stage Variable Selection and Classification Approach for Parkinson's Disease Detection by Using Voice Recording Replications. Comput Methods Programs Biomed 142:147–156. https://doi.org/10.1016/j.cmpb.2017.02.019

Naranjo L, Fuentes-García R, Pérez CJ (2019) A flexible replication-based classification approach for Parkinson's disease detection by using voice recordings. Springer, Cham, pp 81–94. https://doi.org/10.1007/978-3-030-31551-1_7

Nasim AM, Islam MZ (2017a) Forest PA: constructing a decision forest by penalizing attributes used in previous trees. Expert Syst Appl 89:389–403. https://doi.org/10.1016/j.eswa.2017.08.002

Nasim AM, Islam MZ (2017b) ForEx++: a new framework for knowledge discovery from decision Forests. Australas J Inf Syst. https://doi.org/10.3127/ajis.v21i0.1539

Orozco-Arroyave JR, Hönig F, Arias-Londoño JD, Vargas-Bonilla JF, Daqrouq K, Skodda S, Rusz J, Nöth E (2016) Automatic detection of Parkinson's disease in running speech spoken in three different languages. J Acoust Soc Am 139(1):481–500

Ozturk S, Unal Y (2020) A two-stage whale optimization method for classification of Parkinson s disease voice recordings. Int J Intell Syst Appl Eng 8(2):84–93. https://doi.org/10.18201/ijisae.2020261589

Perez CJ, Campos Roca Y, Naranjo L, Martin J (2016) Diagnosis and tracking of Parkinson's Disease By Using Automatically Extracted Acoustic Features. J Alzheimer's Dis Parkinsonism. https://doi.org/10.4172/2161-0460.1000260

Polat K, Nour M (2020) Parkinson disease classification using one against all based data sampling with the acoustic features from the speech signals. Med Hypotheses 140:109678. https://doi.org/10.1016/j.mehy.2020.109678

Postuma RB, Lang AE, Gagnon JF, Pelletier A, Montplaisir JY (2012) How does Parkinsonism start? Prodromal Parkinsonism motor changes in idiopathic REM sleep behaviour disorder. Brain 135(6):1860–1870

Pramanik M, Pradhan R, Nandy P (2021a) Biomarkers for detection of Parkinson's disease using machine learning: a short review. Advances in intelligent systems and computing, 1248th edn. Springer, Berlin, pp 461–475. https://doi.org/10.1007/978-981-15-7394-1_43

Pramanik M, Pradhan R, Nandy P, Bhoi AK, Barsocchi P (2021b) Machine learning methods with decision forests for Parkinson's detection. Appl Sci 11(2):581

Rios-Urrego CD, Vásquez-Correa JC, Vargas-Bonilla JF, Nöth E, Lopera F, Orozco-Arroyave JR (2019) Analysis and evaluation of handwriting in patients with parkinson's disease using kinematic, geometrical, and non-linear features. Comput Methods Programs Biomed 173:43–52

Roffo G, Melzi S (2017) Ranking to learn. Springer, Cham, pp 19–35. https://doi.org/10.1007/978-3-319-61461-8_2

Roffo G, Melzi S, Castellani U, Vinciarelli A (2017) Infinite latent feature selection: a probabilistic latent graph-based ranking approach. Proc IEEE Int Conf Comput vis. https://doi.org/10.1109/ICCV.2017.156

Romexsoft (2017) How to improve medical diagnosis using machine learning. ***2017. https://www.romexsoft.com/blog/improve-medical-diagnosis-using-machine-learning/

Rueda A, Krishnan S (2017) Feature analysis of dysphonia speech for monitoring Parkinson's disease. Annu Int Conf IEEE Eng Med Biol Soc (EMBC). https://doi.org/10.1109/EMBC.2017.8037317

Rusz J, Hlavnička J, Tykalová T, Bušková J, UlmanováEvžen OR, Karel A (2016) Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behaviour disorder. Sleep Med 19:141–147

Sakar BE, Erdem Isenkul M, Okan Sakar C, Sertbas A, Gurgen F, Delil S, Apaydin H, Kursun O (2013) Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Inform 17(4):828–834

Sakar CO, Serbes G, Gunduz A, Tunc HC, Nizam H, Sakar BE, Tutuncu M, Tarkan Aydin M, Isenkul E, Apaydin H (2019) A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. Appl Soft Comput 74:255–263

Smekal Z, Mekyska J, Galaz Z, Mzourek Z, Rektorova I, Faundez-Zanuy M (2015) Analysis of phonation in patients with parkinson's disease using empirical mode decomposition. In: 2015 International Symposium on Signals, Circuits and Systems (ISSCS), pp 1–4

Solana-Lavalle G, Galán-Hernández JC, Rosas-Romero R (2020) Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features. Biocybern Biomed Eng 40(1):505–516

Srivastava T (2019) 11 important model evaluation metrics for machine learning everyone should know. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics. Accessed 11 Feb 2021

Su M, Chuang K-S (2015) Dynamic Feature Selection for Detecting Parkinson's Disease through Voice Signal. In: 2015 IEEE MTT-S 2015 International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-BIO), pp 148–49

Teixeira JP, Oliveira C, Lopes C (2013) Vocal acoustic analysis: jitter, shimmer and HNR parameters. Proced Technol 9:1112–1122. https://doi.org/10.1016/j.protcy.2013.12.124

Thadewald T, Büning H (2007) Jarque–Bera test and its competitors for testing normality–a power comparison. J Appl Stat 34(1):87–105

Tracy JM, Özkanca Y, Atkins DC, Ghomi RH (2020) Investigating voice as a biomarker: deep phenotyping methods for early detection of Parkinson's disease. J Biomed Inform 1(104):103362

Tuncer T, Dogan S (2019) A novel octopus based Parkinson's Disease and gender recognition method using vowels. Appl Acoust 155:75–83

Tuncer T, Dogan S, Acharya UR (2020) Automated detection of Parkinson's disease using minimum average maximum tree and singular value decomposition method with vowels. Biocybern Biomed Eng 40(1):211–220

Upadhya SS, Cheeran AN, Nirmal JH (2019) Discriminating Parkinson diseased and healthy people using modified MFCC filter bank approach. Int J Speech Technol 22(4):1021–1029. https://doi.org/10.1007/s10772-019-09647-0

Viswanathan R, Khojasteh P, Aliahmad B, Arjunan SP, Ragnav S, Kempster P, Wong K, Nagao J, Kumar DK (2018) Efficiency of Voice Features Based on Consonant for Detection of Parkinson's Disease. In: 2018 IEEE Life Sciences Conference (LSC), pp 49–52

Wu K, Zhang D, Guangming Lu, Guo Z (2018) Learning acoustic features to detect Parkinson's disease. Neurocomputing 318:102–108

Xiong Y, Lu Y (2020) Deep feature extraction from the vocal vectors using sparse autoencoders for Parkinson's classification. IEEE Access 8:27821–27830. https://doi.org/10.1109/ACCESS.2020.2968177

Yaman O, Ertam F, Tuncer T (2020) Automated Parkinson's disease recognition based on statistical pooling method using acoustic features. Med Hypotheses 135:109483. https://doi.org/10.1016/j.mehy.2019.109483

Yuvaraj R, Murugappan M, Rajendra Acharya U, Adeli H, Ibrahim NM, Mesquita E (2016) Brain Functional connectivity patterns for emotional state classification in Parkinson's disease patients without dementia. Behav Brain Res 298:248–260

Zeng W, Liu F, Wang Q, Wang Y, Ma L, Zhang Yu (2016) Parkinson's disease classification using gait analysis via deterministic learning. Neurosci Lett 633:268–278

Zhang H-H, Yang L, Liu Y, Wang P, Yin J, Li Y, Qiu M, Zhu X, Yan F (2016) Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples. Biomed Eng Online 15(1):122. https://doi.org/10.1186/s12938-016-0242-6